RESEARCH ARTICLE

Hybrid Deep Learning Algorithms for Dog Breed Identification–A Comparative Analysis

B. VALARMATHI[1], **N. SRINIVASA GUPTA**[2], **G. PRAKASH**[3], **R. HEMADRI REDDY**[4], **S. SARAVANAN**[5], **AND P. SHANMUGASUNDARAM**[6]

[1]Department of Software and Systems Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
[2]Department of Manufacturing Engineering, School of Mechanical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
[3]Department of Database Systems, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
[4]Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
[5]Department of Electronics and Communication Engineering, Srinivasa Ramanujan Centre, SASTRA Deemed University, Kumbakonam, Tamil Nadu 612001, India
[6]Department of Mathematics, College of Natural and Computational Sciences, Mizan-Tepi University, Mizan Teferi 5140, Ethiopia

Corresponding author: P. Shanmugasundaram (psserode@mtu.edu.et)

ABSTRACT Deep learning and computer vision algorithms will be applied to find the breed of the dog from an image. The goal is to have the user submit an image of a dog, and the model will choose one of the 120 breeds stated in the dataset to determine the dog's breed. The proposed work uses various deep learning algorithms like Xception, VGG19, NASNetMobile, EfficientNetV2M, ResNet152V2, Hybrid of Inception &Xception, and Hybrid of EfficientNetV2M, NASNetMobile, Inception &Xception to predict dog breeds. ResNet101, ResNet50, InceptionResNetV2, and Inception-v3 on the Stanford Dogs Standard Datasetswere used in the existing system. The proposed models are considered a hybrid of Inception-v3 &Xception and a hybrid of EfficientNetV2M, NASNetMobile, Inception & Xception. This hybrid model outperforms single models like Xception, VGG19, InceptionV3, ResNet50, and ResNet101.The authors used a transfer learning algorithm with data augmentation to increase their accuracy and achieved a validation accuracy score of 71.63% for ResNet101, 63.78% for ResNet50, 40.72% for InceptionResNetV2, and 34.84% for InceptionV3. This paper compares the proposed algorithms with existing ones like ResNet101, ResNet50, InceptionResNetV2, and InceptionV3. In the existing system, ResNet101 gave the highest accuracy of 71.63%. The proposed algorithms give a validation accuracy score of 91.9% for Xception, 55% for VGG19, 83.47% for NASNetMobile, 89.05% for EfficientNetV2M, 87.38% for ResNet152V2, 92.4% for Hybrid of Inception-v3 &Xception, and 89.00% for Hybrid of EfficientNetV2M, NASNetMobile, Inception &Xception. Among these algorithms, the Hybrid of Inception-v3 &Xception gives the highest accuracy of 92.4%.

INDEX TERMS EfficientNetV2M, hybrid of EfficientNetV2M, hybrid of inception and Xception, NasNet-Mobile, ResNet152V2, VGG19, Xception.

I. INTRODUCTION

Nowadays, there is an increasing demand and usage of image classification and verification techniques; the most significant technique used for image data classification is a deep learning model known as Convolutional Neural Networks (CNN). Deep learning is increasingly becoming a crucial

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

tool in artificial intelligence applications. Optical Character Recognition and facial recognition are two examples of computer vision applications. These areas offer impressive results, fueling increasing interest in deep learning. Classifying images is a field in which deep learning excels. CNN is the most popular deep-learning method for classifying images.

The proposed work is to investigate a variety of Convolutional Neural Network models for classification. The created algorithm might be used in a mobile or online application.

Transfer training, which helps us produce effective outcomes, was used to create the CNN that categorizes dog breeds.

Convolutional Neural Networks (CNNs) are made up of neurons with adjustable biases and weights. The dot product is computed by each neuron using input data. Convolution Neural Network architectures differ from other neural networks in that it relies on actual images as inputs, unlike other neural network types. This enables the inclusion of particular components in the design. CNN lower the number of observable variables that characterize a network. Single-layer neurons act independently from prior layers.

A convolutional neural network architecture from the Inception family called Inception-v3 employs an additional classifier, factorized 7×7 convolutions, and label smoothing to carry label information further down the network.

Deep-separable convolution layers make up the complete architecture of the CNN called Xception.

A type of Convolution Neural Network called NASNet was found through research into neuronal design. Normal cells and decreasing cells serve as the foundation.

EfficientNet is a CNN architecture; all depth, width, and resolution factors are scaled using a compound coefficient method. The EfficientNet scaling technique, in contrast to conventional practice, uniformly adjusts network width, depth, and resolution using a collection of fixed scaling coefficients.

Residual Networks, or ResNets, rather than learning non-referenced functions, learn residual functions about layer inputs. The residual nets enable those layers to match a residual map instead, assuming that each stacked layer exactly matches a desired underlying map. ResNets stack residual blocks together to build networks, such as ResNet-50, which has 50 layers.

A deep convolutional neural network is generally challenging to train from the start. Because size and depth data set relevant to the neural network are rare, using transfer learning as a feature extractor is optimal. A model was already trained on a large data collection. Transfer learning is a crucial component of deep convolutional neural networks that offers solutions to these issues. In computer vision, transfer learning is defined as using a pre-trained model. A pre-trained model must have been trained on a sizable benchmark dataset to handle an issue similar to ours. The magnitude of the new dataset and how closely it resembles the existing dataset play a significant role in determining transfer learning. So, before using transfer learning, a few circumstances shall be considered. The efficacy of the CNN is reduced by overfitting when the new dataset is smaller but contains the same data as the old dataset. If the new dataset is large and has content comparable to the old data, the model can be refined using the entire network.

II. LITERATURE SURVEY

In this research [1], to categorize and identify pet dogs' faces, the researchers offered an improved Yolov3 model. Eight distinct breeds of pet dogs were used to construct the data set for this study. The data set was split into two categories: training and testing, with the training set being used to train the established model. Using more widespread pet dogs as samples, it was suggested by the authors a yolov3-based pet dog categorization model. Instead of using the pet dog's general characteristics, the model chose the face of the pet dog and marked the dog category by detecting facial features, similar to face detection. The trials' findings demonstrated that this approach was capable of rapidly and correctly locating the position of a pet dog's face, and it resolved the issue of pet dog detection and categorization.

In this paper [2], the authors studied how animal identification in veterinary practice was managed through machine learning. Electronic animal health records included digital image graphs using image processing and recognition technology to identify animals. The authors studied how combining "soft" biometrics, such as breed and facial biometrics, could improve dog identification. The researchers applied transfer learning from GoogLeNet to propose Breed-Net for breed classification and subsequently to propose DogNet for identifying individual dogs within the classified breeds.

In this article [3], by determining a dog's breed in a given image, this work aimed to solve the problem of fine-grain and multi-class image recognition. One of the sophisticated deep learning approaches used in the research system was convolutional neural networks. Two distinct networks were constructed and assessed using the Stanford Dogs dataset. Convolutive neural networks' application and evaluation were demonstrated using a software system. It had a central server and a mobile client with resources and tools for online and offline neural network analysis. Two distinct convolutional neural network architectures were presented: the Inception-ResNet-v2 deep architecture and the NASNet-A mobile architecture. Deep Inception-ResNet-v2 model outperformed even the smaller, mobile-friendly CNN, with results that were still encouraging.

This study [4] provided two models for categorizing dogs into different breeds. Due to the increasing difficulty of classifying dogs and the fact that these classifications were based on deep learning, forming the two models that provide different levels of accuracy at both ends requires a fully defined data set. Since every model was periodically subjected to predictions, the researchers encountered numerous functioning levels during the investigation that weren't considered in earlier research. The essential concept of transfer learning, which dealt with the data augmentation technique and its capacity to increase the size of the data set, is also built upon by their approach. Afterward, accuracy levels were matched or compared with both models to establish a comparison for both models. A detailed procedure was also used to classify the data. The comparison between Inception V3 and VGG16 was offered in the publication. Observations showed that Inception V3 offered an accuracy of 85, whereas VGG16 provided a much lower accuracy of 69 than the Inception V3 model.

In this article [5], deep learning made it possible to train algorithms (models) that could categorize and forecast data based on the knowledge that was extracted (learned) from the raw data. Convolutional Neural Networks were one method that was commonly employed for image categorization and detection. In this research, the authors discussed dog breed/type identification after providing a CNN-based method for dog detection in potentially complex images. The findings had a 64% accuracy rate for 120 additional, less common dog varieties and a breed classification accuracy of about 85% for a group of 50 dog breeds. A big data processing infrastructure using a variety of GPUs and an iOS application supporting image classification techniques were used. To enhance the integrity of the data, several preparation methods were employed.

In this study [6], there were many different species and organisms in the world today. This highlights how important it was to classify different tangible objects. Determining the similarities between distinct classes also became extremely important in light of the continuing genetics and evolution study being conducted by scientists worldwide. The experiment that provided the basis for this essay involved classifying different canine breeds using a CNN (Convolutional Neural Network). This algorithm will identify an estimated breed if a canine image is found. The type of dog that most closely resembled a human, if one was given, was identified. The authors created a pipeline to handle photos from the actual world. The dog breed classifier did an excellent job, with extremely good accuracy.

In this article [7], the researchers used cutting-edge models on Imagenet data sets. To extract the feature from the data set used to identify dog breeds, the pre-trained model and learned weights were used. After that, data augmentation and fine-tuning were used to improve the performance of their test's breed categorization accuracy. The performance of the proposed methods was compared with the GoogleNet, DenseNet-169, DenseNet-121, and ResNet-50 models from the most current Image-Net dataset. Their respective test accuracies were 82.08%, 84.01%, 85.37%, and 89.66%, demonstrating the proposed method's superior performance to earlier efforts on Stanford dog breed datasets. Stanford dog breed datasets presented a modified approach to cutting-edge networks like ResNet, DenseNet, and GoogleNet. Due to the limited training datasets, data augmentation and fine-tuning were conducted to improve the test set experiments' accuracy.

In this paper [8], the character of an animal like a dog has changed significantly from earlier generations due to extensive breeding or cross-breeding. Contrary to eye-only recognition, image processing for breed analysis allowed the most accurate prediction of the exact outcome or results. Breed analysis and identification were made using the ADA boosting methodology. By combining numerous weak classifiers, ADA Boosting produced a robust classifier. The authors employed image processing classification to distinguish between the many dog breeds. It accurately predicted the dominant breed or breeds present in the canine. The

proper breed(s) were identified using image processing algorithms for breed classification because the dogs might be cross-breed ancestors. It would be necessary for straightforward breed-based dog categorization, and it could demonstrate how unreliable or trivial breed identification by the naked eye is. The authors could study and perform animal recognition using image processing, including sheep, cattle, and others. VGG Net was an extensive collection of pre-defined CNN demonstrating its ability to utilize/prepare pictures on most subjects accurately.

This article [9] used Convolutional Neural Networks specifically for dog breed detection. Despite being an effective method, convolutional neural network classification still has a few flaws. For training, Convolutional Neural Networks require many images and much time to get better classification accuracy. The authors employed transfer learning to get around this lengthy period. Learning in computer vision refers to training the CNN using a pre-trained model. A pre-trained model was trained to solve a classification problem comparable to the one the researchers had using transfer learning. In this research, the authors trained over 1400 photos encompassing 120 dog breeds using various pre-trained models, including VGG16, Xception, and InceptionV3. Then the bottleneck features were used in these pre-trained models to identify bottlenecks. Last, Logistic Regression, a multiclass classifier, was used to determine the dog breed from the photos. It achieved 91%, 94%, and 95% validation accuracy for the various pre-trained models VGG16, Xception, and InceptionV3.

The dog was one of the most common domesticated animals. Having so many dogs led to several issues, including population control, decreased rabies outbreaks, vaccination control, and formal ownership. Currently, there are about 180 distinct canine breeds. Each canine breed had distinctive characteristics and health problems. To deliver the appropriate therapies and training, it was essential to identify individuals and their breeds. The article illustrates the classification approaches for dog breed classification using two image processing techniques. 1) Two instances of conventional techniques were the Histogram of Oriented Gradient and the Local Binary Pattern (LBP). 2) The deep learning-based strategy utilizing transfer learning and Convolutional Neural Networks (CNN). The outcome demonstrated that their CNN model that had been trained better performed when classifying dog breeds. Compared to 79.25% accuracy when using the HOG descriptor, it achieved 96.75% accuracy [10].

With improved techniques, image classification made significant progress and improved accuracy. However, there was a ton of room for improvement regarding fine-grained classification. The different animals could all be recognized from the image, but it was more challenging to determine the breed of each animal. This essay aims to advance the classification of animal breeds. Several pre-trained deep learning models were trained and tested using the standard Stanford dog breed dataset. The pre-train network was tweaked, and the results were compared. The training process data recorded while

fine-tuning the AlexNet model was displayed in graph and tabular forms. Comparative analysis was performed using each network that has recorded data, and the results are displayed. DenseNet201 reported a testing accuracy of 87.15%, GoogleNet reported a testing accuracy of 81.53%, AlexNet reported a testing accuracy of 84.35%, and ResNet50 reported a testing accuracy of 90.12%. The final three layers of each network were modified to produce these results. The study could be expanded by considering more models and experimenting with layer modifications [11].Classifying images of dogs by breed was difficult, making it a problematic picture classification task. In this study [12], a Convolutional Neural Network (CNN) and transfer learning model-based Android application was created that analyses images to identify a dog's breed. The Android application allows users to capture or upload dog photos. The features required for testing were then extracted after the picture underwent pre-processing. Based on transfer learning and CNN, dog breed predictions were made. The model was taught using Stanford's standard dog dataset, and it had a 94% accuracy rate when tested against actual data. The intended work had been effectively designed, implemented, and tested. The authors had created a simple Android application that allowed users to submit or choose an image to determine a dog's breed. There was a very short waiting time because it worked without an internet link and gave the answer right away. In this study it was shown how to build a dog breed recognition model and deploy it on an Android device using pre-trained models. The application's size could be reduced in the future with the help of enhancements. Additionally, the model that correctly predicts photographs captured in another way could be improved.

Deep learning neural networks have recently gained popularity and are used in various industries, including finance and healthcare, travel, media, retail, etc. The methodology for optimizing CNN, as used in the Stanford dataset of canine breeds, was presented in the current work [13]. The deep neural network, which had weights and biases, was comparable to the Convolution Neural Network. The specific features or patterns contained in the original data were predicted using CNN filters. Modern technologies frequently employed trained Convolutional Neural Networks that had been fine-tuned. Many refined transfer learning techniques are in use now. Inception-ResNet-V2 was implemented over the dataset in this application. Only mentioning the dog's breed was insufficient; it was also crucial to mention its origin, color, height, weight, longevity, health, training, and other traits specific to each breed. Web scraping or web data extraction was used to retrieve data from websites. As a result, web scraping could complete the same activity faster than manually copying data from websites because the process was automated. Reference websites like Wikipedia and Dog Breed List were used in this application to obtain crucial data that was then rendered utilizing web scraping with a respectable user interface and user experience. This paper gave numerous traits and crucial information about the dog based on the supplied image result.

The authors used InceptionV3, MobileNet, VGG- 16, & Xception algorithms, and Stanford Dogs (ST) Dataset, Columbia Dogs (CU) Dataset & Flickr-dog Dataset were used in this paper [14]. CU and ST were used to classify canine breeds, and Flickr-dog was considered for dog identification. The identification rate of dogs was 78.09% without using ''soft'' biometrics, but by using a decision network to combine ''soft'' biometrics, the identification rate could reach an accuracy of 84.94%. The suggested strategy, which relied solely on CNNs, produced average accuracy gains of 6.2%. The identification process using a ''fusion'' on a classifier decision achieved approximately 11.2% higher accuracy.

This paper [15] used Xception and Multilayered Perceptron (MLP) algorithms. The dog breed dataset was derived from a Kaggle contest dog breed identification. The dataset comprised 120 unique dog breeds and 10,222 images of dogs. LogLoss and Balanced Accuracy; the optimal model produced an accuracy of 0.5480, or 54.80%. The results were obtained using only three splits. The accuracy achieved by this model was not satisfying. The number of splits could be raised, allowing the model to train more and improving prediction accuracy. Other approaches may be tried that are more accurate at predicting dog breeds. Additionally, various breed combinations could be trained to detect variations in accuracy.

In this article [16], the authors put forth a brand-new framework model that was referred to as SC-MPEM (Supervised Clustering Using Multi-Part CNN and EM), which makes use of the Inception v3 network for training and the YOLOv3 for discriminative part detection. There were four distinct benchmark datasets used, like the Oxford-IIIT Pet dataset (OD), the Columbia Dogs with Parts (CD), Stanford Dogs (SD), and camera trap pictures from the Snapshot Serengeti datasets. It has been demonstrated that deep CNNs trained under supervision on a sizable and diverse dataset extract superior features than most traditional methods, even for unsupervised tasks. The novel yet straightforward proposed approach outperformed other state-of-the-art models. To further increase the stiffness of the training dataset, images of the animals could be embraced in various poses (facing away from the camera) under various lighting circumstances (day and night). The animal detection algorithms could use thermal images to prevent illumination issues with visible images.

This paper [17] used Convolutional Neural Networks with modals like InceptionResNet V2 and InceptionV3. Also,the researchers used the Stanford Dogs dataset, containing 120 unique dog breeds and 10222 dog images for training, and 10357 images for testing.

This study [18] used a convolutional neural network and TensorFlow model called MobileNet for mobile and embedded mobile application visions. A self-made dataset of 1000 dog images were used. The algorithm used here (CNN) gave good accuracy for all the tested datasets. Transfer learning made an excessive decision by combining a prebuilt model with the model developed in this research. Analysis

was done only for dogs, but planning to extend this project so that other animals can be identified.

In this article [19], Convolutional Neural Networks (CNN) with pre-trained models like ResNet50 were used in this paper [19]. The authors used dog images of 133 different dog breeds. 8,351 dog images and 13,233 human images in total were used to identify dogs only when they were confirmed to be canines, allowing the system to determine whether the given image more resembled a person or the closest predicted dog breed. 82.7% accuracy using CNN with ResNet50 algorithm (transfer learning). Of all the algorithms, it was discovered that this one had the highest accuracy. 12% of dogs were misclassified as people, which could be decreased by using a bigger dataset.

In this paper [20], CNN with three representative models, VGG16, Inception V3, and Xception was used in this work [20]. Stanford Dog data set with 20,580 dog images for 120 different breeds of dogs was used in this work. The accuracy score of Xception was 99%, Inception was 94%, and VGG16 was 85%. One drawback was that learning was too slow; when the dataset grew bigger, VGG16 + LR (Logistic Regression) did not perform as well. The learning rate could be improved by considering other models.

In this study [21], a deep learning-based technique for identifying dog breeds using face photos was demonstrated in the study. To increase accuracy, the suggested approach combines pre-trained CNNs with the transfer learning technique. Three CNN models—MobilenetV2, InceptionV3, and NASNet—were examined in the experiments. Each model was developed using training sets of images enhanced with random noise, rotation, and other effects. With a rotation image training set, the NASNet model achieved the highest accuracy of 89.92%. The rotation might help with picture alignment because the model primarily concentrated on the center of the images. With a classification accuracy of more than 80% in all scenarios, the suggested approach could deliver a promising performance. It may be highly accurate with enhanced datasets like rotation and translation.

This paper [22] classified different dog breeds using CNN. If a canine image were provided, the algorithm would search for the breed of the dog and similarities in the breed's features. If a human image was provided, it was determined which facial features would show in a dog and vice versa.

In this paper [6], the classification of different canine breeds was done using the convolutional neural network. If an image of a dog is discovered, this algorithm will estimate the breed. An associated dog breed was determined if a human image was given. The researchers developed a pipeline for processing real-world photos. This method could be improved by teaching it to distinguish between humans and dogs. Accuracy might be increased even further through data augmentation. The network could identify features independent of orientation or scale due to data augmentation. It is obvious that using transfer learning to create a convolutional neural network was much more accurate than creating one from the start.

In this study [3], the authors offered a fine-grained, multi-class image identification challenge to identify a dog's breed in a given image precisely. The presented system used modern deep learning methods, including Convolutional Neural Networks. Two different networks were trained and assessed using the Stanford Dogs dataset. A software system demonstrated the application and assessment of Convolutional Neural Networks. It included a central server, a mobile client, and parts and modules for online and offline neural network analysis. The Inception-ResNet-v2 deep architecture and the NASNet-A mobile architecture were two different convolutional neural network designs that had been introduced. The designs were evaluated using a particular image classification challenge: identifying dog breeds. The pre-trained networks were adjusted using the Stanford Dogs dataset. The findings were encouraging even for the smaller, mobile-friendly CNN, with only 10% less accurate than the deep Inception-ResNet-v2 mode.

In this paper [2], the authors examined how combining "soft" biometrics, such as canine breed and face biometrics, could enhance dog identification. The proposed BreedNet was used to classify breeds, and the proposed DogNet was used to recognize specific dogs within the classed breeds, using transfer learning from Google LeNet. To categorize dog breeds and then recognize specific dogs using photographs, the "coarse-to-fine technique" and transfer learning were used. The proposed BreedNet's breed categorization accuracy was comparable to the highest outcomes previously reported. Breed categorization first reduces the search space for additional canine identification by identifying the top-k potential breeds given a probe picture of a dog. The BreedNet learned for breed categorization was converted to DogNet for canine identification using transfer learning, allowing the same CNN architecture at both the "coarse" and "fine" stages. Comparing the suggested method to previous works, accuracy was 15% higher.

In this article [23], the authors used Convolutional Neural Networks to categorize dog breeds in a highly exact way. It fell under the domain of fine-grained image classification problems, in which inter-class variances were modest, and one little area of the image analyzed often makes the difference in categorization. ImageNet classes could have considerable inter-class variances, making it easier to categorize accurately. It was intended to train and categorize dog breeds using a Convolutional Neural Network framework. It began by employing CNNs based on the LeNet and GoogLeNet architectures.

In this article [24], the researchers found that the two networks used, VGG-16 and DenseNet201, could identify humanly perceptible patterns when fine-tuned on the Stanford Dogs dataset. Even though there was over-fitting in both networks, the necessary measures were taken to avoid and lessen its effects. Their results were presented and analyzed to show that both networks could still identify patterns despite the over-fitting. The authors examined both networks' response maps (or feature maps) to identify breed-specific

characteristics. Combining their knowledge of networks and 35 features, the experts could interpret the networks given Locke's theory of ideas and words. Although the article's authors rejected the idea that these networks were conscious, they were a good match for a Lockean interpretation.

In this article [25], Multiple Microsoft Kinect v2s were used to record a range of dog breeds, and a motion capture system was also used to acquire the 3D ground truth skeleton. The topic of 3D canine poses estimation from RGBD pictures was the focus of their study. Using this information, several false RGBD images were produced. The authors used prior models of form and pose to constrain a stacked hourglass network trained to forecast the positions of 3D joints. Their model was tested on fictitious and real RGBD images, comparing their findings to previous studies that fitted canine models to images.

This research used a dataset of 70 dog breeds to train and test transfer deep learning algorithms [26]. The dataset was statistically stable, containing approximately 100 images of each category of dog breeds. Then, different deep learning methods were used, such as Convolutional Neural Network, InceptionNet, InceptionResNet, VGG16, ResNet, DenseNet, and etc. The results from algorithm training and testing were compared based on measures like accuracy, precision, recall, and area under the curve.

In this study [27], the authors provided instructions for building a residual neural network to categorize dog breeds according to a sporting category. The system's objective was to make it easier for people to recognize the different canine breeds. The Tsinghua Dogs dataset provided the five distinct dog breed types used. The same setup was used to evaluate ResNet 50 and ResNet 101, two CNN implementations. Based on the study's results, ResNet 101 demonstrated improved macro-average f1-score outcomes while maintaining high accuracy. A ResNet 50 f1-score was 84%, while a ResNet 101 f1-score was 86%.

To classify acute lymphoblastic leukemia (ALL) using microscopic white blood cell images [28], the authors suggested a hybrid Inception v3 XGBoost model for their work. The XGBoost model served as the classification and Inception v3 as the image feature extractor.

III. DATASET DESCRIPTION & SAMPLE DATA
The dataset used for this research is available at the link provided below.

http://vision.stanford.edu/aditya86/ImageNetDogs/

The Stanford Dogs dataset includes images of 120 distinct canine breeds. This dataset was produced using images and annotation from ImageNet for the purpose of fine-grained image categorization. Below is a list of what this dataset contains.
- 120 categories
- 10,222 images
- Class labels and bounding boxes are annotations

Out of 120 categories, the sample five dogs' breed name (Boston bull, dingo, Pekinese, bluetick & Golden Retriever) and the corresponding ID is shown in Table 1.

TABLE 1. Dog breed name with ID.

S. No.	id	breed
1	000bec180eb18c7604dcecc8fe0dba07	boston_bull
2	001513dfcb2ffafc82cccf4d8bbaba97	dingo
3	001cdf01b096e06d78e9e5112d419397	pekinese
4	00214f311d5d2247d5dfe4fe24b2303d	bluetick
5	0021f9ceb3235effd7fcde7f7538ed62	golden_retriever

FIGURE 1. Sample Data from the dataset.

Out of 10,222 images, one sample image is shown in Figure. 1. Twenty images with corresponding labels of the dogs are shown in Figure 2. The 120 categories of 10,222 image class distribution with count value (i.e.) an image count of each dog breed in the dataset are shown in Figure 3. Out of 120 categories of dog breed names with numbers, the first five samples are 0: 'affenpinscher', 1: 'afghan_hound', 2: 'african_hunting_dog',3: 'airedale', 4: 'american_staffordshire_terrier', and 5: 'appenzeller.'

IV. PROPOSED WORK
To conduct a comparison of the accuracy values in this paper, the suggested methods make use of seven distinct algorithms like Xception, VGG19, NASNetMobile, EfficientNetV2M, ResNet152V2, and two hybrid methods [Hybrid of Inception &Xceptionand Hybrid of EfficientNetV2M, NASNetMobile, Inception &Xception] to predict dog breeds. The proposed work will evaluate the seven algorithms and determine the most precise and effective. Existing algorithms such as ResNet101, ResNet50, InceptionResNetV2 and Inception-v3 are used.

Figure 4 illustrates the various steps that will be taken during the execution of this task to produce the desired result.

Step 1 (Import Modules): The first step is to import important libraries required for the proposed work. The matplotlib and seaborn libraries are used for graphs. scikit_learn library is used for training and testing splitter. Numpy and pandas'

FIGURE 2. A few dog images with corresponding labels.

libraries are used to handle image arrays. Tensorflow is used to utilize the pre-trained models and train our own model.

Step 2 (Load Dataset): Load the Standford dataset in the training folder into the RAM during the runtime.

Step 3 (Analyze and Visualize the Dataset): Analyze the total number of images and the class distribution of the dogs to identify any gaps in the dataset. Visualizing data by using Pandas data frame & functions and plotting graphs.

Step 4 (Validate Dataset): Check if the number of labels and images is equal. If yes, all images are labeled, and it can be proceeded further.

Step 5 (Encode Categorical Classes): Encode categorical classes by assign a unique number to each of the class of the dog breeds.

Step 6 (Convertintoarray): To train a deep learning model using the characteristics of an image, it is necessary to convert an image into an array. The Python NumPy library can

transform images into arrays because it can be mainly used to work with arrays.

Step 7 (Import the Optimizer): The proposed work uses Adam as an optimizer. The adaptive moment estimation provides the foundation for the name Adam. The network weights are updated during the exercise using this optimization algorithm, another extension of stochastic gradient descent. The Adam Optimizer continuously updates the learning rate for each unique network weight. The Adam optimizer is commonly used because of its numerous advantages. The algorithm also runs quicker, requires less memory, and requires less tuning than earlier optimization algorithms. It is also simpler to implement.

Step 8 (Extract Features): Now, the features of the images must be extracted in the form of the NumPy array that was previously framed. For this, each pre-trained model's pre-processor will be used to pre-process the images. After
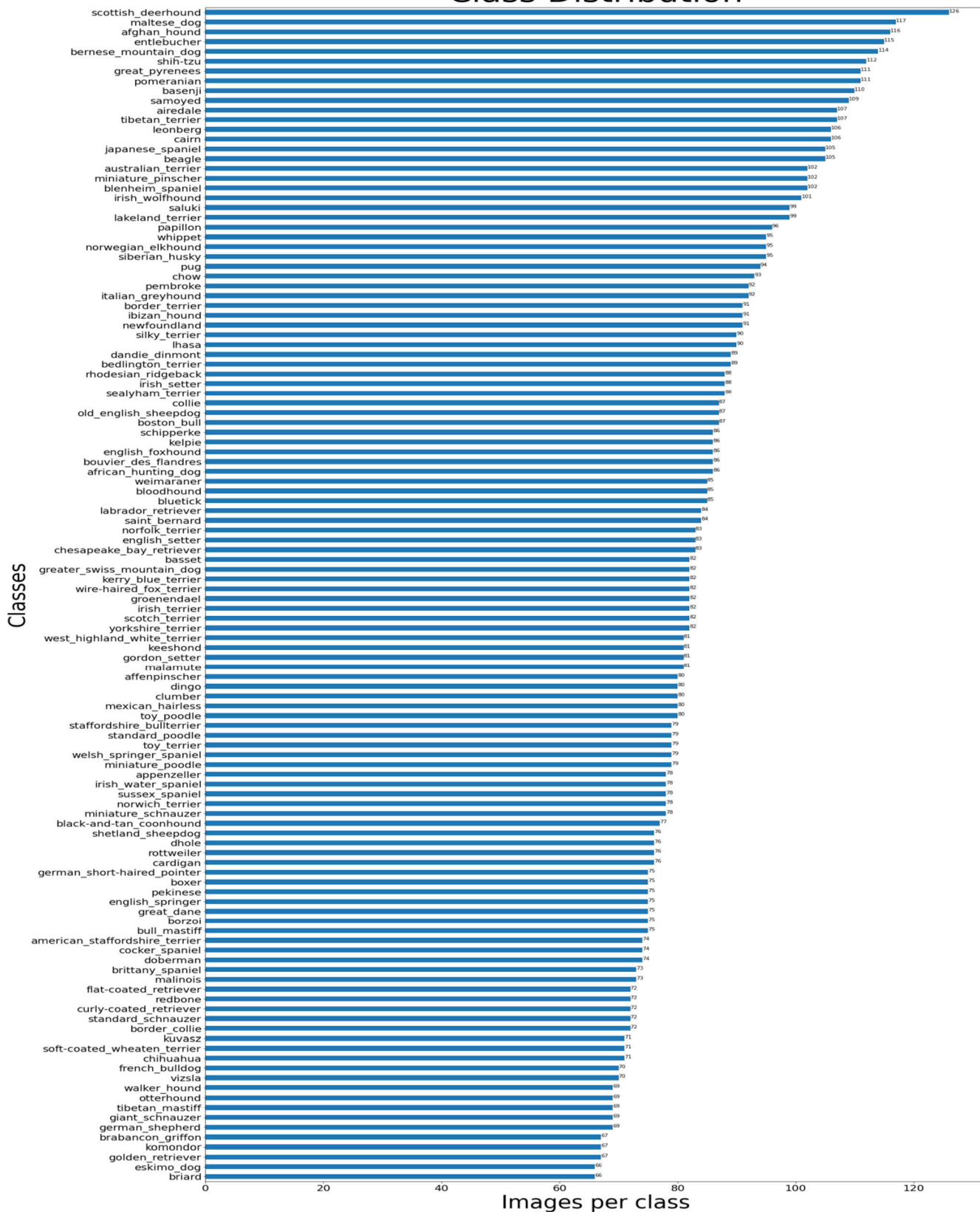
FIGURE 3. The class distribution of dog breed images per class.
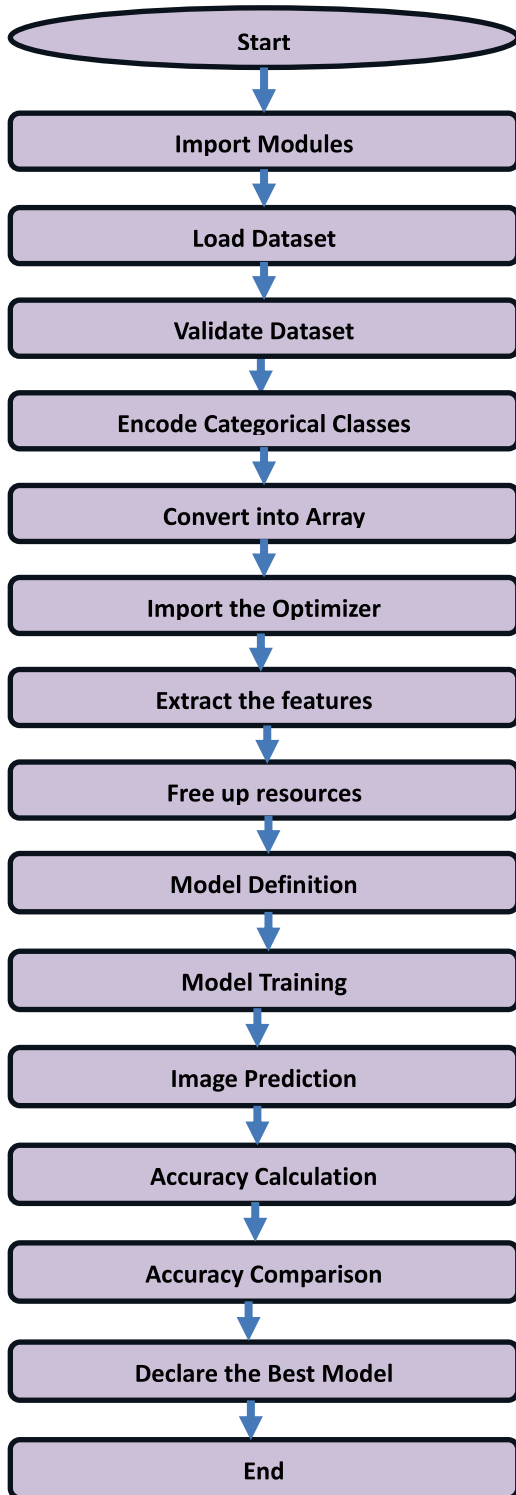
FIGURE 4. Flowchart of the proposed method.

this, feature selection was performed using the models. For the individual model pipeline, it has been used Efficient-NetV2M, NASNetMobile, and ResNet152V2. For the first hybrid model, feature selection was performed using two models (Inception-v3 and Xception); for the second hybrid

model, feature selection was performed using four models (EfficientNetV2M, NASNetMobile, Inception-v3, and Xception) concurrently and then concatenate the extracted features into a single NumPy array.

Step 9 (Free Up Resources): Garbage values were collected to free up some RAM after the feature selection is done.

Step 10 (Model Definition): Now, the proposed model is defined, including the dropout layer and dense layer with SoftMax activation function. The model with Adam Optimizer was finally compiled

Step 11 (MODEL TRAINING): After that, fit the model (train) with the extracted features over 100 epochs and 256 batch size.

Step 12 (Image Prediction): Now, any input image can be used and extract features from the image and predict using the trained model.

Step 13 (Accuracy Calculation): The metric used for the proposed work is accuracy. In this work, it has been calculated training and validation accuracies for the various models.

Step 14 (Accuracy Comparison): Here, the accuracies of various models like Xception, VGG19, NASNetMobile, EfficientNetV2M, ResNet152V2, Hybrid of Inception & Xception, and Hybrid of EfficientNetV2M, NASNetMobile, Inception & Xception are compared.

Step 15 (Declare the Best Model): The next stage is to name the best model based on the highest accuracy achieved after comparing the accuracy of various models.

V. RESULTS AND DISCUSSION

In the existing system [17], the authors achieved a validation accuracy score of 71.63% for ResNet101, 63.78% for ResNet50, 40.72% for InceptionResNetV2, and 34.84% for Inception-v3. Table 2 also lists the results of the training and validation accuracy values for different models in the existing systems. Among these four algorithms, the Inception-v3 algorithm achieved a low validation accuracy value of 34.84%, whereasthe ResNet101 algorithmmachieveda high validation accuracy value of 71.63%. The comparative analyses of the training and validation accuracies for various models in the existing systems are shown in Figure 5.

The proposed algorithms give a validation accuracy score of 91.9% for Xception, 55% for VGG19, 83.47% for NASNetMobile, 89.05% for EfficientNetV2M, 87.38% for ResNet152V2, 92.4% for Hybrid of Inception-v3 & Xception, and 89.00% for Hybrid of EfficientNetV2M, NASNet-Mobile, Inception & Xception. Among these algorithms, the Hybrid of Inception-v3 & Xception gives the highest accuracy of 92.4%.

For the comparative analysis, the proposed model can be trained with Inception and Xception alone. When trained, an accuracy of 91.4% is given with the Inception-v3 and an accuracy of 91.9% with the Xception model. The hybrid model of Inception-v3 and Xception predicted 92.4% accuracy, greater than that of the Inception and Xception models found individually. The least accurate model, with an
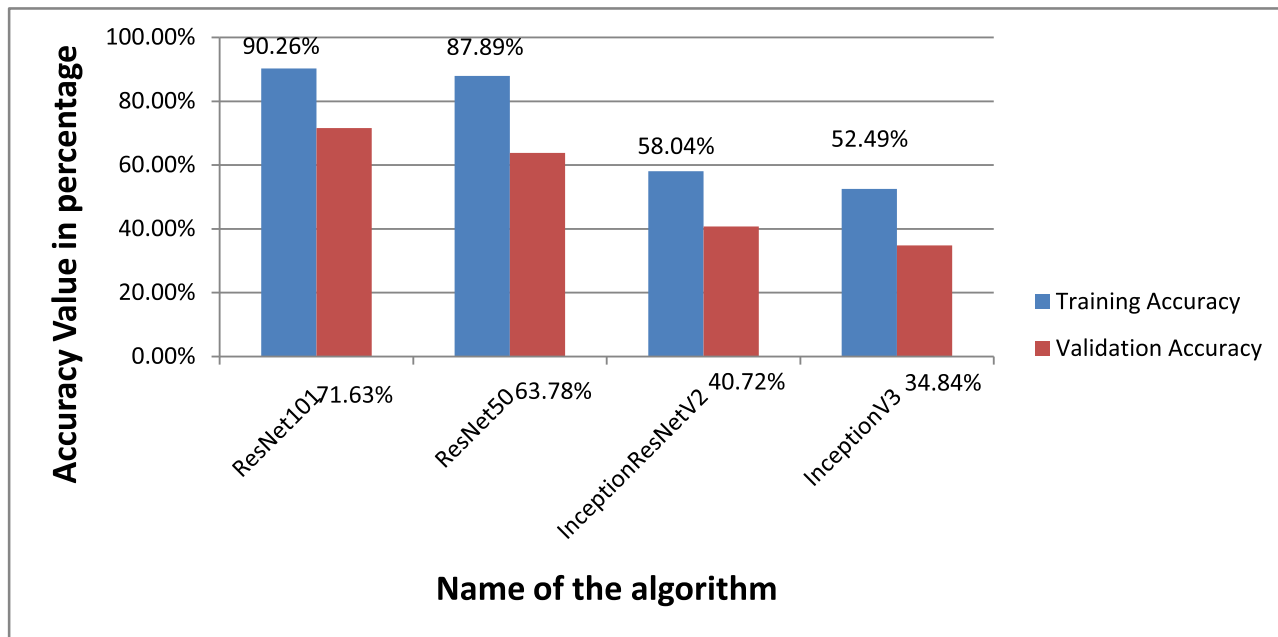
FIGURE 5. Comparative analysis of the training and validation accuracies for various models in the existing systems.
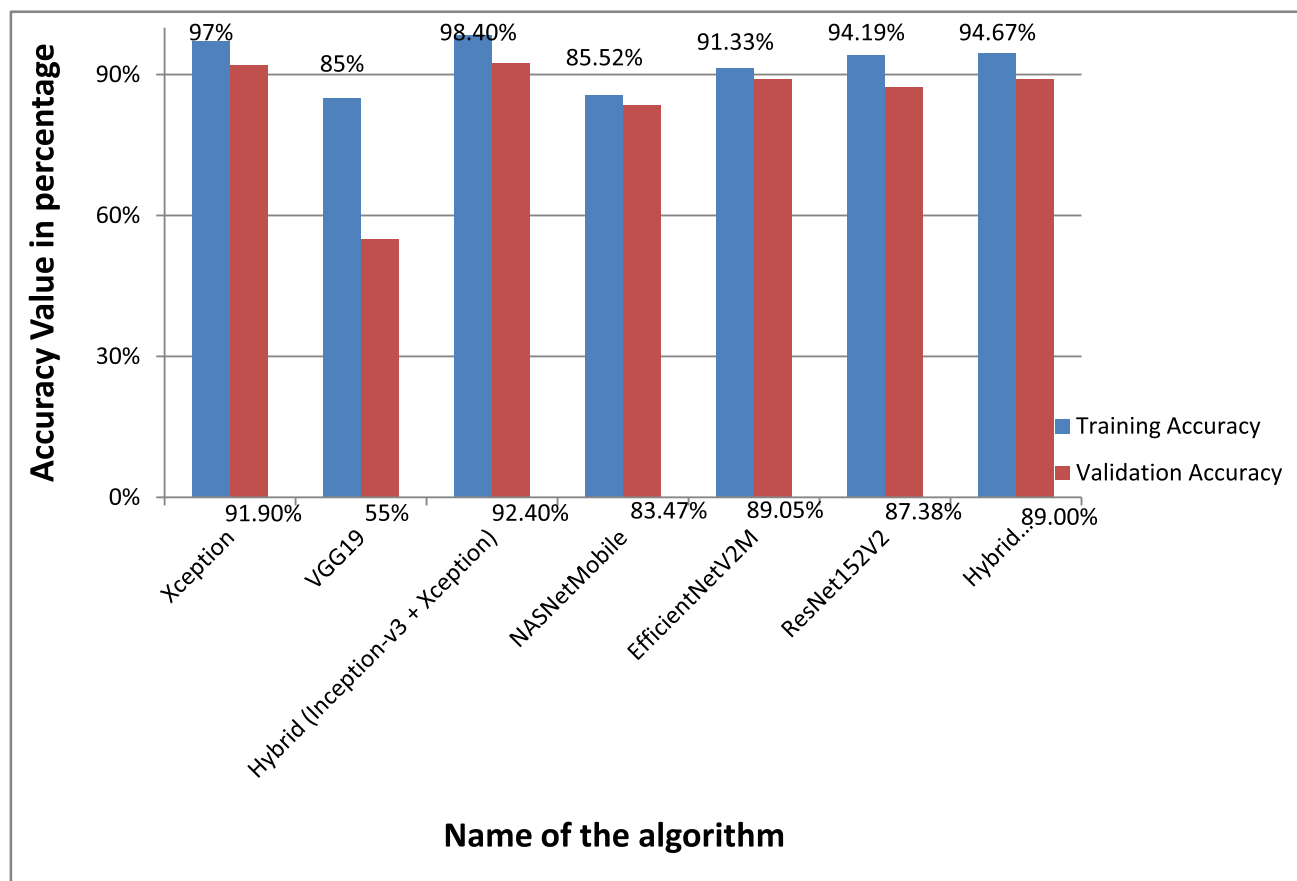

FIGURE 6. Comparative analysis of the training and validation accuracies for various models in the proposed system.

accuracy of 55%, was the VGG19. Table. 3 displays the accuracy values from training and validation for the different models in the proposed methodology.

The comparative analysis of the accuracy of different algorithms during training and validation in the proposed system is shown in the figure 6. In the proposed hybrid model, the

TABLE 2. Training and validation accuracies for various models in the existing systems.

Model	Training Accuracy	Validation Accuracy
Xception	97%	91.9%
VGG19	85%	55%
Hybrid (Inception-v3 + Xception)	98.4%	92.4%
NASNetMobile	85.52%	83.47%
EfficientNetV2M	91.33%	89.05%
ResNet152V2	94.19%	87.38%
Hybrid (EfficientNetV2M+NASNetMobile +Inception+Xception)	94.67%	89.00%

TABLE 3. Proposed methodology accuracy values from training values and validation for the different models.

Model	Training Accuracy	Validation Accuracy
ResNet101	90.26%	71.63%
ResNet50	87.89%	63.78%
InceptionResNetV2	58.04%	40.72%
InceptionV3	52.49%	34.84%

images are taken as input which is converted to an array then it is passed to neural network. Train images are used to train our model. The training accuracy which we got is 98.4% and the validation accuracy which we got is 92.4% for Hybrid(Inception-v3 +Xception). The training accuracy value for VGG19 algorithm which we got is 85% and the validation accuracy which we got is 55%.

VI. CONCLUSION AND DISCUSSION

Considering the large number of breeds in this fine-grained classification problem, we view our overall findings as successful. Given the high variability both between and within the 120 different breeds included in the dataset, we successfully predict the right breed over 92.4% of the time in a single guess, a result that very few models could match. The training accuracy which we got is 98.4% and the validation accuracy which we got is 92.4% for Hybrid (Inception-v3 + Xception). The training accuracy value for VGG19 algorithm which we got is 85% and the validation accuracy which we got is 55%. The proposed algorithms give a validation accuracy score of 91.9% for Xception, 55% for VGG19, 83.47% for NASNetMobile, 89.05% for EfficientNetV2M, 87.38% for ResNet152V2, 92.4% for Hybrid of Inception-v3 & Xception, and 89.00% for Hybrid of EfficientNetV2M, NASNetMobile, Inception & Xception. Among these algorithms, Hybrid of Inception-v3 & Xception gives the highest accuracy of 92.4%.

Other machine learning models, neural networks, and deep learning models should all be further investigated in future

work on canine breed prediction. This is an approach that has promise for upcoming tasks given the success of our hybrid detection network.

Due to time and CPU limitations, it was found difficult to conduct many iterations of our technique using Neural Networks because training each and every layer of the model is a very time-consuming process. So, in order to train the complete model and increase accuracy, we anticipate receiving better GPU and equipment.

We advise further research into neural networks for key point detection, particularly by training networks with different architectures and batch iterators to see what strategies might be more effective. In the end, neural network architectures take a long period to train and iterate, which should be taken into account for future work. However, compared to more conventional methods, neural networks are powerful classifiers and will improve prediction accuracy.

REFERENCES

[1] C. Wang, J. Wang, Q. Du, and X. Yang, "Dog breed classification based on deep learning," in *Proc. 13th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2020, pp. 209–212, doi: 10.1109/ISCID51228.2020.00053.

[2] X. Tu, K. Lai, and S. Yanushkevich, "Transfer learning on convolutional neural networks for dog identification," in *Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2018, pp. 357–360, doi: 10.1109/ICSESS.2018.8663718.

[3] Z. Ráduly, C. Sulyok, Z. Vadászi, and A. Zölde, "Dog breed identification using deep learning," in *Proc. IEEE 16th Int. Symp. Intell. Syst. Informat. (SISY)*, Sep. 2018, pp. 271–276, doi: 10.1109/sisy.2018.8524715.

[4] A. Varshney, A. Katiyar, A. K. Singh, and S. S. Chauhan, "Dog breed classification using deep learning," in *Proc. Int. Conf. Intell. Technol. (CONIT)*, Jun. 2021, pp. 1–5, doi: 10.1109/conit51480.2021.9498338.

[5] R. O. Sinnott, F. Wu, and W. Chen, "A mobile application for dog breed detection and recognition based on deep learning," in *Proc. IEEE/ACM 5th Int. Conf. Big Data Comput. Appl. Technol. (BDCAT)*, Dec. 2018, pp. 87–96, doi: 10.1109/BDCAT.2018.00019.

[6] M. V. S. Rishita and T. A. Harris, "Dog breed classifier using convolutional neural networks," in *Proc. Int. Conf. Netw., Embedded Wireless Syst. (ICNEWS)*, Dec. 2018, pp. 1–7, doi: 10.1109/ICNEWS.2018.8903980.

[7] A. Ayanzadeh and S. Vahidnia, "Modified deep neural networks for dog breeds identification," 2018, Art. no. 2018120232, doi: 10.20944/preprints201812.0232.v1.

[8] A. Sharma, A. Sahoo, M. Azhagiri, and D. Dutta, "Recognition of dog breeds using convolutional neural network and visual geometry group," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 3898–3902, Oct. 2019.

[9] B. V. Kumar and K. Bhavya, "Dog breed identification with fine tuning of pre-trained models," *Int. J. Recent Technol. Eng. (IJRTE)*, vol. 8, nos. 2–11, pp. 3677–3680, Sep. 2019.

[10] P. Borwarnginn, K. Thongkanchorn, S. Kanchanapreechakorn, and W. Kusakunniran, "Breakthrough conventional based approach for dog breed classification using CNN with transfer learning," in *Proc. 11th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2019, pp. 1–5, doi: 10.1109/iciteed.2019.8929955.

[11] D. Kamdar, "Implementation of pre-trained deep learning model for dog breed classification," *Turkish J. Comput. Math. Educ.*, vol. 120, no. 7, pp. 555–558, 2021.

[12] D. D. D. Bhavani, M. H. S. Quadri, and Y. R. Reddy, "Dog breed identification using convolutional neural networks on Android," *CVR J. Sci. Technol.*, vol. 17, no. 1, pp. 62–66, Dec. 2019, doi: 10.32377/cvrjst1711.

[13] S. M. Mohamed, "Dog breed identification using convolution neural network and web scraping," *Int. J. Sci. Res.*, vol. 9, no. 3, pp. 488–491, Mar. 2020, doi: 10.21275/SR20307193747.

[14] K. Lai, X. Tu, and S. Yanushkevich, "Dog identification using soft biometrics and neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8851971.

[15] K. Mulligan and P. Rivas, "Dog breed identification with a neural network over learned representations from the Xception CNN architecture," in *Proc. 21st Int. Conf. Artif. Intell. (ICAI)*, 2019, pp. 1–6.

[16] D. M. Sundaram and A. Loganathan, "A new supervised clustering framework using multi discriminative parts and expectation–maximization approach for a fine-grained animal breed classification (SC-MPEM)," *Neural Process. Lett.*, vol. 52, no. 1, pp. 727–766, Aug. 2020.

[17] G. Kanika, G. Sharankhala, A. Bhavya, and G. Riya, "Dog breed identification using CNN architecture," *Int. J. Sci. Eng. Res.*, vol. 11, no. 6, pp. 1040–1045, 2020.

[18] I. Sneha, "Dog breed prediction using convolutional," *Neural Netw. Int. J. Eng. Adv. Technol.*, vol. 9, no. 5, pp. 318–322, Jun. 2020.

[19] T. Sadanandana, "Dog breed identification using pre-trained models," *International Journal of Computer Science and Mobile Computing*, vol. 9, no. 4, pp. 76–82, 2020.

[20] Z. Wang, D. Zhao, and K. Hong, *Dog Breed Classification*. San Diego, CA, USA: ECE Project Univ. California, 2020.

[21] P. Borwarnginn, W. Kusakunniran, S. Karnjanapreechakorn, and K. Thongkanchorn, "Knowing your dog breed: Identifying a dog breed with deep learning," *Int. J. Autom. Comput.*, vol. 18, no. 1, pp. 45–54, Feb. 2021, doi: 10.1007/s11633-020-1261-0.

[22] B. K. Shah, A. Kumar, and A. Kumar, "Dog breed classifier for facial recognition using convolutional neural networks," in *Proc. 3rd Int. Conf. Intell. Sustain. Syst. (ICISS)*, Dec. 2020, pp. 508–513, doi: 10.1109/ICISS49785.2020.9315871.

[23] D. Hsu, *Using Convolutional Neural Networks to Classify Dog Breeds*, document CS231n, Convolutional Neural Networks for Visual Recognition, Feb. 2015.

[24] X. S. Higa, "Dog breed classification using convolutional neural networks: Interpreted through a lockean perspective," Senior Theses Student Publication, 2019. [Online]. Available: https://core.ac.uk/download/pdf/214315107.pdf

[25] S. Kearney, W. Li, M. Parsons, K. I. Kim, and D. Cosker, "RGBD-dog: Predicting canine pose from RGBD sensors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8333–8342, doi: 10.1109/cvpr42600.2020.00836.

[26] A. K. Agarwal, V. Kiran, R. K. Jindal, D. Chaudhary, and R. G. Tiwari, "Optimized transfer learning for dog breed classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 1s, pp. 18–22, Oct. 2022.

[27] N. H. Pratama, E. Rachmawati, and G. Kosala, "Classification of dog breeds from sporting groups using convolutional neural network," *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, vol. 7, no. 4, pp. 1080–1087, Nov. 2022.

[28] S. Ramaneswaran, K. Srinivasan, P. M. D. R. Vincent, and C.-Y. Chang, "Hybrid inception v3 XGBoost model for acute lymphoblastic leukemia classification," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–10, Jul. 2021, doi: 10.1155/2021/2577375.

B. VALARMATHI was born in Tirukovilur, Tamil Nadu, India. She received the degree in electronics and communication engineering and the master's degree in computer science and engineering from IIT Madras, India, and the Ph.D. degree in information and communication engineering from Anna University, India. She has three decades of teaching, research, and administrative experience. She is currently a Professor of information technology with the Vellore Institute of Technology, Vellore, India. She has published 47 research articles in data mining, machine learning, sentiment analysis, natural language processing, text mining, the Internet of Things, data science, soft computing, and heuristics. She is a Life Member of the ISTE and the Soft Computing Research Society.

N. SRINIVASA GUPTA was born in Tiruvannamalai, Tamil Nadu, India. He received the degree in mechanical engineering and the master's degree in industrial management from IIT Madras, India, and the Ph.D. degree in mechanical engineering from the Vellore Institute of Technology, Vellore, Tamil Nadu, India. He has three decades of teaching, research, and administrative experience. He is currently a Professor of mechanical engineering with the Vellore Institute of Technology. He has published 25 research articles in cellular manufacturing, heuristics, data mining, sentiment analysis, natural language processing, the Internet of Things, data science, soft computing, and text mining. He is a Life Member of ISTE.

G. PRAKASH received the B.E. degree in computer science and engineering from Madras University, in 1995, the M.E. degree in computer science and engineering from Annamalai University, in 2004, and the Ph.D. degree in information and communication engineering from Anna University, in 2015. He is currently an Associate Professor with the School of Computer Science and Engineering, Vellore Institute of Technology, Vellore. He has published 18 articles in Scopus and SCI-indexed journals and 16 papers in Springer and IEEE Xplore-sponsored international conferences. His research interests include information security, agile-based software engineering, cryptography, and steganography. He is a Life Member of various professional bodies, such as ISTE, IAENG, and the Internet Society. He received various awards and certifications for his remarkable contributions to cyber security.

R. HEMADRI REDDY received the Ph.D. degree in mathematics, in 2007. He has ten years of teaching and research experience. He published more than 50 research articles in various reputed international journals. His research interests include biofluid dynamics and machine learning.

S. SARAVANAN is currently an Assistant Professor with the Department of Electronics and Communication Engineering, Srinivasa Ramanujan Centre (SRC), SASTRA Deemed University, Kumbakonam, Tamil Nadu. Before his recent appointment with SASTRA Deemed University, he was an Associate Professor with CMIT, Bengaluru. He has more than 20 years of experience in both teaching and research. So far, he has published more than 70 Scopus and more than five SCI/SCIE-indexed research articles in national and international journals. His research interests include VLSI design, hardware security, machine learning, and embedded systems. He is a Life Member of ISSE and ISTE.

P. SHANMUGASUNDARAM received the Ph.D. degree from Anna University. He is currently an Associate Professor with the Department of Mathematics, College of Natural and Computational Sciences, Mizan-Tepi University, Tepi Campus, Ethiopia. His research title was "Applications of Intuitionist Fuzzy Sets in Decision Making Problems." He has more than 26 years of experience in teaching and more than 15 years of experience in research. He has also been a peer reviewer of various Ph.D. theses, international journals, and conferences. He has published more than 25 Scopus/SCI-indexed research articles. His research interests include fuzzy logic, machine learning, operations research, and research methodology. He is a life member of two professional bodies.

● ● ●