

Comparative Analysis of Image Classification Methods on Cat Breeds and Behavior Using Machine Learning Techniques

Adeel Ansari ¹, Seema Ansari ^{2*}, Sara Syed Prasla ¹, Abeera Naveed ¹

¹ Department of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan

² Department of Electrical Engineering, Institute of Business Management, Karachi, Pakistan

*Corresponding Author: seema.ansari@iobm.edu.pk

Abstract:

This research investigates image classification techniques applied to two distinct datasets related to cats. The primary focus is on addressing the problem of accurately classifying cat breeds and cat behavior. This research focuses on the comparative analysis of both deep learning and machine learning techniques. The techniques are categorized as transfer learning on deep learning models, Transfer learning on machine learning algorithms, and Teachable machine pre-trained models. Transfer learning has gained popularity as one of the techniques employed in the inception of V3 for classifying images. It requires re-utilizing an existing model for a new model by applying a small-scale dataset to pace up training and enhance overall performance. Five different methodologies are explored: Convolutional Neural Networks using the Google Inception-V3 model, Convolutional Neural Networks on top of the Google Inception-V3 model with K-fold cross-validation, Random Forest on Inception V3 features, Support Vector Machine on Inception V3 features, and Teachable Machine model. The study aims to compare the performance of these methodologies in terms of accuracy, F1 score, and ROC-AUC score. The research results show that each approach has levels of effectiveness, with various algorithms and models showing accuracy and F1 scores in classifying both cat breeds and behaviors. These findings offer information on image classification in datasets related to cats, helping to improve the precision of identifying cat breeds and behaviors.

Keywords: Convolutional Neural Networks, F1 Score, Inception V3, K-fold cross-validation, Random Forest, ROC-AUC Score Support Vector Machine, Transfer Learning.

I. INTRODUCTION

Image recognition is an aspect that allows machines to comprehend and interpret visual information in fields. It entails sorting images into categories or tags based on their content. The importance of image recognition has grown alongside the rapid rise of digital images and the demand for automated analysis and decision-making processes. In the realm of feline studies, image recognition can aid in addressing concerns linked to cat breeds and behavior. Cats, being cherished companions, often encounter challenges due to people's knowledge and understanding. Many individuals find it difficult to identify cat breeds, which can affect their care and dietary needs. Moreover, grasping cat behavior is crucial for establishing a nurturing environment for these creatures [1].

Image recognition methods present an opportunity to close this information gap by identifying and analyzing cat breeds and their behavior tendencies. Nevertheless, the current image classification techniques for cats come with a set of challenges and areas that need improvement. To start with, distinguishing between cat breeds can be tricky due to their visual distinctions [2, 3]. Moreover, the wide range of behaviors exhibited by cats makes it difficult to capture and interpret their mood-related cues from images. The lack of annotated datasets specifically focused on cat breeds and behavior only adds to the complexity [4].

Overcoming these hurdles necessitates creating and assessing image classification models designed for tasks related to cats. This study seeks to address the shortcomings in existing methods by exploring approaches, such as those based on learning utilizing cross-validation techniques and making use of preexisting models. The objective is to deepen our knowledge of cat breeds and behaviors ultimately leading to better care practices, treatment options and responsible pet ownership.

Image recognition is a field in the study of computer vision and machine learning. It involves teaching a model to identify and distinguish classes or types of images based on their extracted features [5]. Recent advancements in image recognition include the use of algorithms like automated random forest, image segmentation, and object-based methods. These innovations improve the accuracy and efficiency of classifying kinds of images, including those related to cat breeds and behaviors [6]. The focus of this study is on identifying cat breeds and behaviors using image recognition techniques.

This research aims to investigate the performance of deep learning and machine learning models in the classification of cat breeds and their behaviors. For this, we will assess Convolutional Neural Networks (CNNs) using the Inception-V3 architecture, also with CNNs built on top of Inception-V3 with the incorporation of K-fold cross-validation, a Random Forest classifier leveraging Inception-V3 features, a Support Vector Machine model also utilizing Inception-V3 features and the Teachable Machine model, a cutting-edge machine learning platform. By evaluating the performance of these methods, based on accuracy, F1 score and ROC AUC score, we aim to provide a comprehensive understanding of the capabilities and limitations of each model in accurately identifying the unique characteristics of different cat breeds and their associated behaviors [1], [3].

The goal of this research is to create an image classifier for identifying cat breeds and behaviors by developing and implementing image classification models while also investigating the practical application of these techniques.

This paper reviews the challenges and constraints in methods for categorizing cat breeds and behavior in images and suggests ways to enhance the accuracy and dependability of these models. Additionally, insights and suggestions will be provided on how image classification models can be applied in real-life situations like pet care and behavior assessment to better comprehend and handle cats. This study focuses on assessing image classification models for recognizing cat breeds and behaviors, with the intended audience being researchers, pet owners, veterinarians, and animal welfare organizations.

The objective is to investigate techniques, such as learning methods, k-fold cross-validation, and SVM algorithms to precisely classify cat breeds as well as detect emotions or moods from cat pictures. The study encompasses gathering a dataset of cat images, preprocessing them appropriately, applying chosen techniques using the Inception V3 architecture, and evaluating performance using metrics. The results obtained from this research could advance our knowledge about cats' behavior patterns, enhance care practices, and potentially facilitate the creation of automated systems for analyzing cat behaviors.

The novelty of this research is to use image classification methods for the identification of various cat breeds and their behavior and also present a comparative study of Convolutional Neural Networks using the Google Inception-V3 model, Convolutional Neural Networks on top of Google Inception-V3 model with K-fold cross-validation, Random Forest on Inception V3 features, Support Vector Machine on Inception V3 features, and Teachable Machine model.

A. Hypothesis:

Null Hypothesis (H_0): There is no significant difference in the classification performance between the CNN model and traditional machine learning models (Random Forest and SVM) for cat's behavior and breed classification.

Alternate Hypothesis (H_1): There is a significant difference in the classification performance between the CNN model and traditional machine learning models (Random Forest and SVM) for cat behavior and breed classification.

II. LITERATURE REVIEW

Image classification has become an essential area of research in computer vision, with applications ranging from object recognition to medical diagnosis. In recent years, researchers have turned their attention to the classification of cat images, focusing on identifying specific breeds and behaviors. This literature review synthesizes findings from studies

conducted over various classification models, their challenges and limitations as well as in which application domains have, they been applied at in relevance.

In [2], the authors implied the CNN model for classification over Fruits-360 data sets on Kaggle, using VGG16, VGG19, Inception V3 and DenseNet classifiers. The model converted the output to probabilities through Softmax activation function and with the use of the Adam optimizer, significant losses were minimized. From [2], it was concluded that the Inception V-3 classifier with the CNN model generated a better classification in comparison to the other classifier techniques.

In [6], various image classification techniques for classifying different types of images from different sensors and resolutions were implied. The process included determining a suitable classification system, feature extraction, selecting good training samples, pre-processing and selecting appropriate classification methods, and post-classification, processing, and assessing overall accuracy. Techniques addressed included the maximum likelihood, support vector machine, artificial neural networks, decision tree classifiers, fuzzy-set classifiers and spectral mixture analysis. However, there were limitations in the results due to difficulty in classifying combined SAR and optical images due to a high number of bands, small input feature space, and limited training samples, which led to poor estimates and incorrect generalization.

In [8], a comparative analysis of SVM and CNN for image classification using MNIST and COREL1000 datasets was implied, considering the performance metrics of accuracy and overall computational time. Due to the limitations of the data sets and evaluation metrics, the performance metrics were not successfully able to provide any significant differences between the performance yield of the two techniques.

In [9], the author uses Pytorch to create a convolutional neural network model for animal image classification. The model achieves a high accuracy of above 90%, classifying cats and dogs from wildlife pictures but failed to decipher cats from dogs and vice versa.

The authors in [10] focused on identifying the breed of cats and dogs from images, using a new dataset that covers 37 different breeds. The task was challenging due to the deformable nature of pets and the subtle differences between their breeds. The model classifies cat and dog breeds automatically by capturing the shape of the animals' faces and their fur appearance. They also compared the two classification approaches: hierarchical and flat. Overall results obtained showed a higher accuracy close to 90% with the CNN model.

The authors in [11], have focused on the adoption rates of stray dogs and cats, which pose a huge threat to human communities. Therefore, an image classification approach is considered that will analyze the photo traits to speed up the adoption process and boost the adoption rates, for the animal shelters. The algorithms favorably performed well with higher accuracy scores using the neural network approach.

In [12], the authors developed a semi-supervised learning-based Multi-part Convolutional Neural Network (MP-CNN) that classified animal images into different species. The experimental results proved that with the coalesced approach of MP-CNN and with pseudo-labels, it was able to classify the animal breeds accurately with a score of 99.5%, but only at a broader category of species within the limited data sets.

In [13], the authors developed a robust learning method for animal classification from camera-trapped images with high-noise labels. Two different network structures were considered, with and without clean samples for noise handling. K-means clustering was used to divide the training samples into groups to train different networks. The performance of the method was evaluated using Snapshot Serengeti and Panama-Netherlands datasets. The output was an adequate yield from the technique, but due to the nature of the datasets and messy labeling of noise, the performance of this technique suffered subjectively due to the integrity of the data sets.

After reviewing the relevant literature, we selected the CNN with Inception V3 model (incorporating k-fold cross validation), SVM, Random Forest, and Teachable Machine model for our study on classifying cat breeds and behavior using image analysis.

III. PROBLEM STATEMENT

The problem addressed in this research is the need for accurate and efficient classification of cat behavior based on image data. Current approaches often lack the capability to capture subtle nuances and variations in cat behavior,

leading to misclassification and inadequate understanding of their needs. This knowledge gap can result in mistreatment and suboptimal care for cats. Therefore, there is a need for a robust and reliable image classification system that can accurately identify and differentiate various cat behaviors, enabling better understanding and appropriate responses to their needs.

IV. METHODOLOGY

In this section, we discuss the proposed methodology for this research in the following phases:

B. Phase 1 – Design & Decide

In order for the research to be conducted successfully, we needed to make sure that the datasets needed to perform image classification on cat breeds and cat behaviors were collected and ready for further processing. Moreover, the research serves the purpose of differentiating between the different techniques to see which one works superior to the others. One surprising bend would be that the size of both of the datasets will vary vastly. The dataset for cat breed classification can be acquired through Kaggle.com. however, the dataset for the cat’s emotion detection has to be collected. The assembly process of images into a correct class can be accomplished through web scraping using Flickr API. The image dataset for the cat’s breed has 52 distinct courses and a total of 32,856 images. Nonetheless, the image dataset of the cat’s behavior, as mentioned, has 4 classes and a total of 3,735 images. The next step moving further to classification is a pre-processing and cleaning of images.

C. Phase 2 – Understanding & Evaluating

The research is analytical based which aims to discover which method works best on image classification. The 5 components used are Google Inception-V3, Convolutional Neural Networks (CNN), Random Forest Algorithm, Support Vector Machine (SVM), and Teachable Machine Model [14].

The Inception: V3 model is a widely used pre-trained CNN architecture that has shown excellent performance in various image recognition tasks. The Inception V3 has obtained an accuracy greater than 78.1% over the ImageNet data set. This model has been a combination of various research studies over the years. This model comprises symmetric and asymmetric building blocks, including average pooling, max pooling, convolutions, dropouts, concatenations, and fully connected layers, as illustrated in Figure 1 [16]. Batch normalization is used extensively throughout the model and applied to activation inputs. Loss is computed using SoftMax.

Convolutional Neural Networks (CNN): it performs computations using convolutional layers, which involve sliding a set of filters (kernels) across the input image to extract features as shown in Figure 2 [17]. Each filter performs a dot product between its weights and a local receptive field of the input image. These dot products are summed and passed through an activation function, namely, ReLU, to introduce non-linearity.

The outputs of the convolutional layers are then flattened and fed into one or more fully connected (dense) layers. These layers perform matrix multiplications and apply activation functions to produce the final classification output [17].

Random Forest: it is an ensemble learning algorithm that consists of multiple decision trees as shown in Figure 3. Each decision tree is trained on a randomly selected subset of the data and a random subset of features. At each node of a decision tree, the algorithm evaluates different splitting criteria, such as Gini impurity or information gain, to determine the best feature and threshold for splitting the data. The decision tree grows recursively until a stopping criterion is met, such as reaching a maximum depth or a minimum number of samples per leaf. During inference, each decision tree in the forest independently predicts the class of a given input, and the final prediction is determined by majority voting or averaging the predictions of all the trees [18].

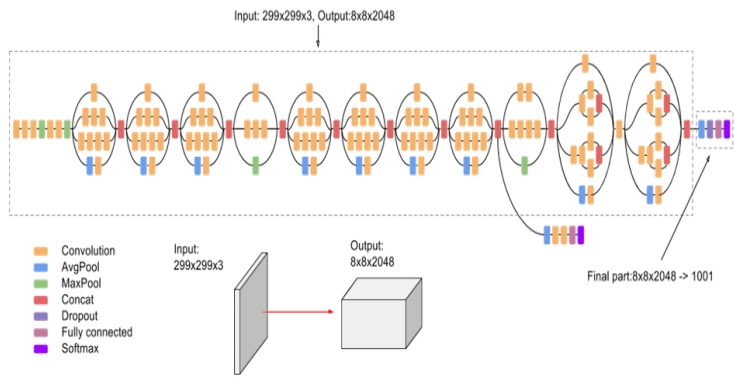


Figure 1: The Inception V3 Model

Source: [16]

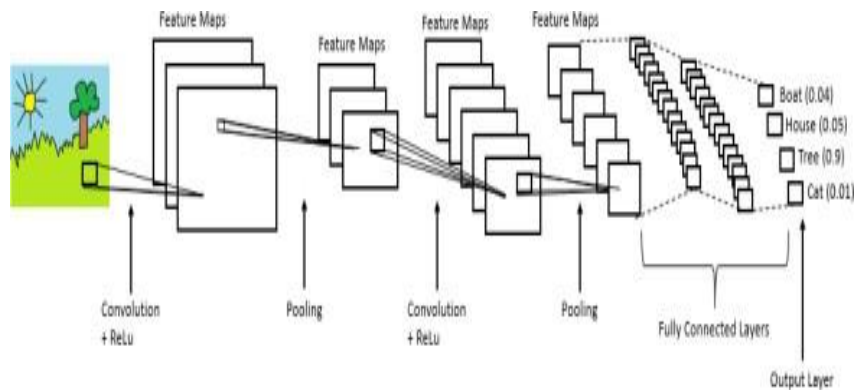


Figure 2: Architecture of Convolutional Neural Networks

Source: [17]

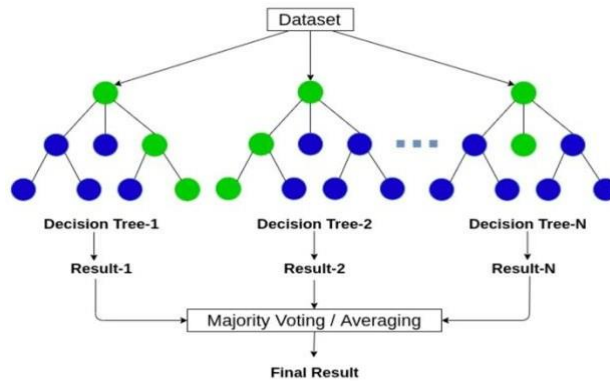


Figure 3: Architecture of Random Forest using multiple decision trees

Source: [18]

Support Vector Machines (SVM): It aims to find an optimal hyperplane that separates the input data points into different classes. The algorithm finds the hyperplane with the maximum margin, which is the distance between the hyperplane and the nearest data points of each class, as shown in Figure 4 [19]. SVMs use a kernel function to map the input data into a higher-dimensional feature space where it can be linearly separable. In this feature space, the algorithm finds the optimal hyperplane using optimization techniques, such as quadratic programming. During inference, the SVM classifies new data points based on which side of the hyperplane they fall.

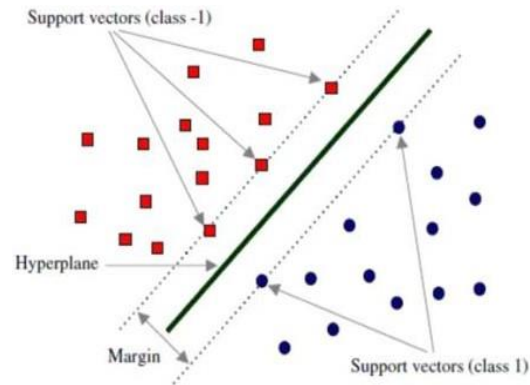


Figure 4: Architecture of Support Vector Machines Source: [19]

Teachable Machine: It is a web-based tool that simplifies the process of training machine learning models. The mathematical calculations, in Teachable Machine can differ based on the type of model architecture employed, like a convolutional neural network or a pre-trained model. Nevertheless, the tool offers a to use interface for users to categorize and educate their models with labelled data. The calculations required for training a Teachable Machine model involve optimization methods to adjust the models' weights using labeled data and loss functions to assess the models' effectiveness.

The dataset will be utilized in such a way, as to capture the diverse range of cat breeds and their associated behaviors. The dataset will be preprocessed and organized to ensure consistent formatting and resolution across all images.

To provide input data for the various deep learning and machine learning techniques, we will apply the following approaches:

- **Convolutional Neural Networks:** It uses the Google Inception-V3 model. The cat images will be directly fed into the Inception-V3 architecture, allowing the model to extract the relevant visual features for classification.
- **Convolutional Neural Networks:** It is on top of the Google Inception-V3 model with K-fold cross-validation: The Inception-V3 model will be used as a feature extractor, and the extracted features will be used as inputs to the CNN models. K-fold cross-validation will be employed to improve the performance.
- **Random Forest on Inception-V3 features:** The Inception-V3 model will be used to extract features from the cat images, and these features will then be used as inputs to the Random Forest classifier.
- **Support Vector Machine on Inception-V3 features:** Similar to the Random Forest approach, the Inception-V3 features will be used as inputs to the Support Vector Machine model for classification.
- **Teachable Machine model:** The Teachable Machine platform will be used to directly train on the cat images, allowing for an efficient and user-friendly classification solution.

By leveraging these diverse input data strategies, we aim to thoroughly evaluate the performance of each deep learning and machine learning technique, providing significant insight into the most effective approaches for classifying cat breeds and behaviors.

D. Phase 3 – Recommended Practices

In the last stage of our methodology, we chose these methods because they are known to be effective in tackling the research problem of image classification.

Convolutional Neural Networks (CNNs) have demonstrated a better performance in image classification tasks by learning hierarchical features from images. However, the Inception-V3 model, even though it is a widely used pre-trained CNN architecture, has demonstrated a far better performance in different image recognition experiments. Combining these two methods is expected to outperform others. To enhance this method, we incorporate K-fold-cross validation. This technique helps evaluate the models' performance by testing them on subsets of the dataset and providing a thorough assessment of their efficacy. The Inception V3 model serves as a foundation for both Random Forest and SVM models. By using features extracted from the Inception V3 model to train a Random Forest classifier, we can leverage learning and decision trees to classify images based on learned patterns. Additionally, applying an SVM classifier atop the Inception V3 model helps extract features

Teachable Machine enables users to develop image classification models without requiring programming expertise. Through the use of these techniques, the study seeks to investigate how well they can classify images and determine the appropriate method, for the datasets, at hand.

V. EXPERIMENTAL SETUP

The setup, for the experiment involved utilizing Anaconda Navigator and Jupyter Notebook as the software platform to execute the five methods. A gaming laptop with 32GB of RAM and a GPU was used for hardware support. Anaconda Navigator offered a user interface for managing the Python packages and dependencies, while Jupyter Notebook enabled coding development and execution. The decision to utilize a laptop equipped with a GPU was crucial for enhancing the efficiency of training and inference processes in learning models, significantly speeding up tasks. This configuration ensured that the experiments were carried out on hardware and software setups to ease the implementation and assessment of image classification techniques.

All 5 procedures are performed on both datasets. All procedures have same data preparations, data augmentation process to keep it simplified and so the base of all methods does not over-complicate the results and analysis based upon them.

To evaluate the performance of the five deep learning and machine learning techniques, we will follow a structured workflow encompassing data preparation, model training, and output generation.

A. Data Preparation

- Gather the dataset of cat images covering a wide range of breeds and behaviors.
- Preprocess the images to ensure consistent size, resolution, and format across the dataset.
- Split the dataset into training, validation, and testing subsets to enable model evaluation and tuning.

B. Convolutional Neural Networks (CNNs) using Google Inception-V3 model

- Utilize the pre-trained Inception-V3 model as the backbone of the CNN architecture.
- Fine-tune the Inception-V3 model on the cat image dataset, allowing the model to learn the distinctive visual features of different cat breeds and behaviors.
- Train the CNN model end-to-end, optimizing the model parameters for accurate classification.
- Evaluate the model's performance on the testing dataset and report the classification accuracy.

C. CNNs on top of Google Inception-V3 model with K-fold cross-validation

- Use the Inception-V3 model as a feature extractor, obtaining the deep visual features for each cat image.
- Divide the dataset into K folds and train K different CNN models, each using the Inception-V3 features as inputs.
- Employ K-fold cross-validation to ensure robust and reliable model performance, mitigating the risk of overfitting.
- Evaluate the average classification accuracy across the K models and report the results.

D. Random Forest on Inception-V3 features

- Leverage the Inception-V3 model to extract visual features from the cat images.
- Train a Random Forest classifier using the Inception-V3 features as input variables.
- Optimize the hyperparameters of the Random Forest model through techniques like grid search or cross-validation.

- Evaluate the Random Forest model's performance on the testing dataset and report the classification accuracy.

E. Support Vector Machine on Inception-V3 features

- Similar to the Random Forest approach, use the Inception-V3 features as inputs to a Support Vector Machine (SVM) classifier.
- Tune the hyperparameters of the SVM model, such as the kernel function and regularization parameter, to achieve optimal performance.
- Assess the SVM model's classification accuracy on the testing dataset and compare the results with the other techniques.

F. Teachable Machine model:

- Leverage the Teachable Machine platform, which allows for efficient and user-friendly model training and deployment.
- Directly input the cat images into the Teachable Machine and train the model to classify the breeds and behaviors.
- Evaluate the Teachable Machine model's performance on the testing dataset and report the classification accuracy.

Overall, several preprocessing steps were applied to the datasets, including data augmentation, normalization, and data splitting. These processing steps were the same in all 5 different methods used on distinct datasets. It was performed in order to simplify the results so that each method is based on the unified processing of data. Data augmentation techniques such as rotation, width and height shifting, and horizontal flipping were employed using the 'ImageDataGenerator' from the 'tensorflow.keras.preprocessing.image' module. These techniques help to increase the diversity and variability of the training data, leading to improved model generalization.

Image normalization was conducted using settings within the 'ImageDataGenerator.'

- The 'rescale' feature of the 'ImageDataGenerator' adjusted the values of the images to fall within a range of 0 to 1.
- This step aims to ensure numerical ranges across all input images, which can aid in refining the training process and model convergence.
- With the 'rotation_range' parameter, images are randomly rotated by up to 30 degrees. This technique helps the model adapt better to variations in object orientations.
- The 'width_shift_range' parameter introduces shifts in images by a maximum of 20% of the image width. This adjustment mimics changes in object positions within the image.
- Similarly, the 'height_shift_range' parameter applies shifts in images by a maximum of 20% of the image height. This variation simulates shifts in object positions within the image.
- By utilizing the 'horizontal_flip' setting, images are randomly flipped horizontally, adding diversity to the dataset through mirrored objects.

The dataset is divided into training, validation and testing sets using the 'train_test_split' function. The dataset consisted of 32,856 cat images, covering a diverse range of breeds and behaviors. Data Partitioning:

- The dataset was split into training, validation, and testing subsets using a 70/15/15 ratio, respectively.
- The training set contained 23,000 cat images, the validation set had 4,928 images, and the testing set had 4,928 images. This partitioning allows for effective model training, hyperparameter tuning, and unbiased evaluation of the model's performance on unseen data. Sample Data:
- The cat images in the dataset varied in resolution, with the majority being high-quality photographs in the range of 800x600 to 1200x900 pixels.
- The dataset included images of common cat breeds, such as Persian, Siamese, Maine Coon, and Bengal, as well as less common breeds and mixed-breed cats.

- In addition to breed information, the dataset also included annotations for various cat behaviors, including playing, sleeping, grooming, and hunting. By providing these details on the dataset size, partitioning, and sample characteristics, we aim to give a comprehensive understanding of the experimental setup, which is crucial for ensuring the reproducibility and validity of the research findings.

By applying data augmentation and normalization, the models become more robust to variations in the input data and are better able to learn meaningful patterns. These preprocessing steps contribute to improving the accuracy and performance of the image classification models during training and evaluation.

All models except the teachable machine modelling process include Inception V3 as a base and transfer learning model. Below is how it is incorporated into the process.

```
'i_model=InceptionV3(weights='imagenet', include_top=False, input_shape=(299, 299, 3))'
```

 initializes an instance of the InceptionV3 model pre-trained on the ImageNet dataset. The `'include_top=False'` argument indicates that the fully connected layers at the top of the network should be excluded, and the `'input_shape=(299, 299, 3)'` specifies the input size of the images.

The `'for layer in i_model.layers: layer.trainable = False'` loop freezes the weights of all the layers in the InceptionV3 model, making them non-trainable. This ensures that the pre-trained weights are kept fixed during the training process.

Among 4 of these processes, two of our methods use the same training process. Those two are CNN on top of the Inception V3 model and CNN on top of Inception V3 with K-fold Cross Validation.

Both processes have the same Convolutional Neural Network layers set up for further process:

- `'model.add(i_model)'` adds the pre-trained InceptionV3 model to the sequential model. This serves as the feature extractor for the subsequent layers.
- `'model.add(GlobalAveragePooling2D())'` adds a global average pooling layer, which reduces the spatial dimensions of the feature maps and extracts their spatial information.
- `'model.add(Dense(128))'` adds a fully connected layer with 128 units.
- `'model.add(Dropout(0.2))'` adds a dropout layer that randomly sets a fraction of input units to 0 at each update during training. This helps to prevent overfitting.
- `'model.add(Dense(51, activation='softmax'))'` adds the output layer with 51 units and uses the softmax activation function to obtain class probabilities.
- `'model.summary()'` displays a summary of the complete model architecture, including the number of trainable parameters and the output shape of each layer.

The model is compiled with the SGD optimizer and categorical cross-entropy loss function. The `'fit'` function is used to train the model on the training data `'x_train'` and validate it on the validation data `'x_val'`. The number of training steps per epoch `'steps_per_epoch'` and validation steps `'validation_steps'` are specified. The training is performed for a specified number of epochs.

The only exception in these processes is the amount of K- fold cross-validation incorporated in order to avoid overfitting. The K-fold cross-validation is performed using the `'Kfold'` function set to 5 from the `'sklearn.model_selection'` module. The dataset is divided into K folds, and for each fold, the model is trained on the training data and validated on the validation data.

The Random Forest Algorithm on the Inception V3 pre-trained model is applied to the features extracted from the InceptionV3 model for image classification. An instance of the `'RandomForestClassifier'` from `'sklearn.ensemble'` is created with 100 estimators (decision trees). The Random Forest classifier is then fitted to the training features and labels using the `'fit'` function. The trained Random Forest classifier is used to make predictions on the validation features. The accuracy of the predictions is calculated by comparing them to the true labels. The classifiers' performance is assessed using a confusion matrix.

Inception V3 features are utilized by the SVM model with extracted features serving as input. A Support Vector Classifier (SVC) instance, from `'sklearn.svm'` is instantiated. The SVC model is trained on the input features and labels using the `'fit'` function. Subsequently, predictions are made on validation data using the trained SVM classifier and accuracy is determined by comparing these predictions with labels. The confusion matrix is also computed to evaluate the performance of the classifier. Teachable Machine model is the pre-trained model that is loaded from the saved model file using `'tf.keras.models.load_model()'`. The loaded model is compiled with the optimizer set to `'adam'`, the

loss function set to 'categorical_crossentropy', and the metric set to 'accuracy'. The preprocessed test set ('x_test_preprocessed') is fed into the trained model to obtain predictions. The predictions are converted to class labels using 'np.argmax()'. Accuracy, confusion matrix, F1 score, and ROC-AUC score are calculated to evaluate the performance of the model on the test set. Overall, the training procedures involve fitting the models to the training data and validating them on separate validation data. The evaluation procedures involve using the trained models to make predictions on test data or validation data and assessing their performance using appropriate metrics such as accuracy, F1 score, and confusion matrix.

VI. EXPERIMENTS AND RESULTS

The results below are of the Cat's breed classification. Table 1 shows the overall performance scores of all five models used for the classification of the different cats' breeds. However, the results from the experiments indicate that the performance of the machine learning models in accurately classifying cat behavior was limited. The accuracy scores, precision, recall, F1 scores, and ROC-AUC scores achieved by the tested methods were relatively low, suggesting limited success in accurately classifying the images [11].

Furthermore, we have observed that different algorithms and models yielded unpredictable results in terms of accuracy and F1 scores for both breed and behavior classification, as shown in Table 1 and Table 2. The cat behaviors that were identified were fearfulness, activity/ playfulness, aggression, sociability, excessive grooming and litterbox issues. However, due to the differences in the breeds, the genetic background of each feline's personality differed significantly. The results are also illustrated in a graphical representation, as shown in Figure 5 and Figure 6. Accuracy can be calculated as the ratio of correctly classified samples to the total number of samples, while the F1 score is a measure that combines precision and recall to provide an overall evaluation of the model's performance. ROC-AUC score is a measure of the model's ability to distinguish between classes and is often used in classification problems. Considering the performance metrics, the researchers found that the CNN with Inception V3 model performed better than other machine learning models.

The results from the different models pertaining to the classification of cat's behaviorism are as shown in Table 2. Based on the results, it was observed that the CNN with Inception V3 model (incorporating k-fold cross-validation) performed the best among the tested machine learning models, achieving an accuracy score of 46.40%. It was also found that the Random Forest and Support Vector Machine models had lower accuracy scores of 23.65% and 30.83%, respectively, while the Teachable Machine model achieved an accuracy score of 37.57%. Therefore, it was hypothesized that the deep learning algorithm Convolutional Neural Network would outperform other machine learning models in classifying cat breeds and behaviors. Also, greater precision values have been achieved for CNN with Inception V3(K fold cross validation), with 30.43% in Table 1 and 45.11% in Table 2.

Table 1: Results of Different Models in terms of Breed Classification

Cats' Breed Classification						
ID	EXPERIMENTS	ACCURACY	PRECISION	RECALL	F1 SCORE	ROC-AUC SCORE
1	CNN with Inception V3	33.75%	29.13%	9.23%	28.23%	84.57%
2	CNN with Inception V3 (K-fold Cross Validation)	34.68%	30.43%	11.03%	29%	85%
3	Random Forest with Inception V3	22.07%	20.18%	7.83%	19.46%	48.15%
4	Support Vector Machine with Inception V3	24.77%	25.23%	8.45%	19.68%	53.88%
5	Teachable Machine Model	13.34%	11.78%	2.83%	10.74%	53.26%

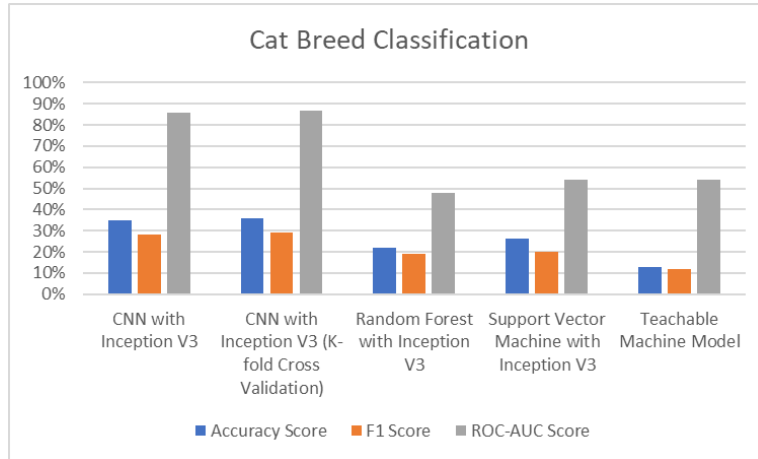


Figure 5: Cat Breed Classification Performance Scores for various models

Table 2: Results of Different Models in terms of Breed Classification

Cats' Behavior Classification						
ID	EXPERIMENTS	ACCURACY	PRECISION	RECALL	F1 SCORE	ROC-AUC SCORE
1	CNN with Inception V3	44.31%	33.1%	9.01%	42.59%	67.62%
2	CNN with Inception V3 (K-fold Cross Validation)	46.40%	45.11%	12.11%	46.24%	72.85%
3	Random Forest with Inception V3	23.65%	24.12%	12.90%	21.26%	46.67%
4	Support Vector Machine with Inception V3	30.83%	29.01%	15.65%	23.77%	51.36%
5	Teachable Machine Model	37.57%	34.55%	12.44%	36.97%	64.98%

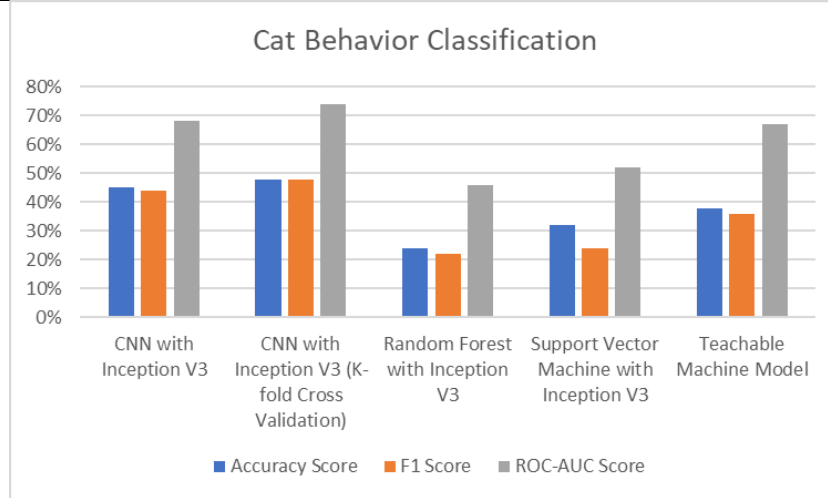


Figure 6: Cat Behavior Classification Performance Scores for various models

The results showed that the CNN with Inception V3 (K fold cross validation) outperformed the other methods, achieving the highest F1 score of 29% in Table 1 and 46.24% in Table 2, following the same with the ROC-AUC score with the highest of 85% and 72.85% in terms of cat breeds and behavior classification.

VII. DISCUSSION

Based on the experimental findings shown in Table 1 and Table 2, the CNN with Inception V3 model using K-fold Cross-Validation achieved the highest ROC-AUC score of 85% and 72%, cat breed and cat behavior classification, respectively. This method worked superior among the five methods tested. The reason why ROC-AUC is better is because of its flexibility when dealing with imbalanced class distribution datasets, just like the ones we use. ROC-AUC takes into account both true positive rate (sensitivity) and false positive rate, providing a more comprehensive assessment of model performance across different class distributions. It is computed based on the ranking of predicted probabilities, rather than a specific threshold. ROC-AUC is less affected by class imbalance and misclassification costs. It focuses on the overall performance of the model in distinguishing between classes, making it more reliable than F1 score and test accuracy.

There are several limitations and challenges encountered in this study. Firstly, the low accuracy and F1 scores suggest that the models struggled to accurately classify the images in both datasets. This may be due to the complexity and variability of cat breeds and behaviors, as well as the limited amount of training data available. Additionally, the use of pre-trained models like Inception V3 may not be optimized for the specific classification task at hand, leading to suboptimal results.

The hypothesis stated that deep learning algorithm CNN would perform better than machine learning models. In like manner, the experimental results strongly support this hypothesis, as all the methods achieved relatively low accuracy, precision, recall and F1 scores, then Convolutional Neural Network. This suggests that other factors, such as the dataset quality, model architecture, and training parameters, may have contributed to the overall performance. Several factors could explain the observed outcomes of other models' low accuracy, precision, recall and F1 scores. Firstly, the limited size of the datasets may have restricted the models' ability to learn complex patterns and generalize well to unseen data. Additionally, the use of generic pre-trained models like Inception V3 may not have captured the specific features and characteristics relevant to cat breed and behavior classification. Fine-tuning or training models from scratch using larger and more representative datasets could potentially improve performance. Overall, while some of the models did not meet the expectations, they provide valuable insights into the challenges of image classification on cat datasets. Further research and experimentation are necessary to overcome the limitations identified and develop more accurate and robust models for cat breed and behavior classification.

VIII. CONCLUSION

The main findings of the research indicate that the task of image classification for cat breeds and behaviors is challenging. The accuracy scores, precision, recall, F1 scores, and ROC-AUC scores achieved by the tested methods were relatively low, suggesting limited success in accurately classifying the images. Among the methods, the CNN with Inception V3 model performed the best but still fell short of achieving satisfactory results.

The proposed solution, which involved applying different methods such as CNN, Random Forest, SVM, and Teachable Machine, aimed to address the problem of cat breed and behavior classification. However, the results indicate that the proposed machine learning models did not effectively tackle the problem, as the models struggled to achieve better accuracy, precision, recall and F1 scores. Nonetheless, the deep learning algorithm, i.e. Convolutional Neural Network with the ROC-AUC score gave far better results than any other model and metric system. Therefore, the proposed solution can be considered partially effective at best.

For future research, exploring alternative deep learning architectures specifically designed for fine-grained image classification of cat breeds and behaviors could yield better results. Collecting a larger and more diverse dataset, potentially incorporating additional features like cat body language or audio data, could also enhance the models' performance. Additionally, investigating advanced data augmentation techniques and employing transfer learning methods tailored to the unique characteristics of cat images could be valuable areas of exploration.

ACKNOWLEDGMENT

I would like to express my deepest gratitude to Dr. Seema Ansari, and to the respective students, namely, Abeera Naveed, and Syed Sara Prasla for their invaluable contribution and dedication to this research on machine learning to identify cat breeds and behavior. Their insights and support were crucial in the successful completion of this project.

CONFLICT OF INTEREST

There is no conflict of interest between all the authors.

REFERENCES

- [1] J. Wang, S. Mall, and L. Perez, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning".
- [2] Tripathi, Milan. "Analysis of convolutional neural network based image classification techniques." *Journal of Innovative Image Processing (JIIP)* 3.02 (2021): 100-117.
- [3] K. Szyc. "Comparison of Different Deep-Learning Methods for Image Classification". Jun. 2018. 10.1109/ines.2018.8523931.
- [4] O. M. Parkhi, A. Vedaldi, A. Zisserman and C. V. Jawahar. "Cats and dogs". Jun. 2012. 10.1109/cvpr.2012.6248092.
- [5] N. Manohar, S. Subrahmanya, R. Bharathi, Y. H. S. Kumar and H. K. G. "Recognition and classification of animals based on texture features through parallel computing". Aug. 2016. 10.1109/ccip.2016.7802872.
- [6] Y. Wang and Z. Wang. "A survey of recent work on fine-grained image classification techniques". *Journal of Visual Communication and Image Representation*. vol. 59. pp. 210-214. Feb. 2019. 10.1016/j.jvcir.2018.12.049.
- [7] Nath, Siddhartha Sankar, et al. "A survey of image classification methods and techniques." 2014 International conference on control, instrumentation, communication and computational technologies (ICCICCT). IEEE, 2014
- [8] Wang, P. En F., and Peng W., "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning." *Pattern Recognition Letters* 141: 61-67, 2021.
- [9] Qinhan, Y., "Animal Image Classifier Based on Convolutional Neural Network", SHS web of conferences, 2022.
- [10] Tighare, K., Bhandekar, R., Ragite, H., 'Image classification between two animals', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 373–376, 2021, doi:10.32628/cseit2174103.
- [11] Pradeepa J., et al. "A review of techniques for image classification to enhance online animal adoption speed", *Proceedings of the 2020 12th International Conference on Computer and Automation Engineering, ACM Other conferences*, 2020.
- [12] Divya, S., Meena, L., Agilandeswari, "An Efficient Framework for Animal Breeds Classification Using Semi-Supervised Learning and Multi-Part Convolutional Neural Network (MP-CNN)", *IEEE Access*, 2019.
- [13] Ahmed A., Hayder, Y., Roland, K., Zhihai, H., "Animal species classification using deep neural networks with noise labels", *Ecological Informatics*, 2020.
- [14] Chutani, P., et al. "High-End Image Classification on the Dogs vs. Cats Dataset Using Convolutional Neural Network." *International Scientific and Practical Conference "Technology in Agriculture, Energy and Ecology"* (TAEE2022), 2022.
- [15] Bai C., Huang L., Pan X., Zheng J., Chen S., "Optimization of deep convolutional neural network for large scale image retrieval", *Neurocomputing*, 2018 Aug 16;303:60-7.
- [16] Singh T, Vishwakarma DK, "A deeply coupled ConvNet for human activity recognition using dynamic and RGB images", *Neural Computing and Applications*. 2021 Jan;33:469-85.
- [17] Phung VH, Rhee EJ, "A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets", *Applied Sciences*. 2019 Oct 23;9(21):4500.
- [18] Kibria HB, Matin A, "The severity prediction of the binary and multi-class cardiovascular disease– A machine learning-based fusion approach", *Computational Biology and Chemistry*. 2022 Jun 1;98:107672.
- [19] García-Gonzalo E, Fernández-Muñiz Z, Garcia Nieto PJ, Bernardo Sánchez A, Menéndez Fernández M, "Hard-rock stability analysis for span design in entry-type excavations with learning classifiers", *Materials*. 2016 Jun 29;9(7):531.
- [20] Carney M, Webster B, Alvarado I, Phillips K, Howell N, Griffith J, Jongejan J, Pitaru A, Chen A, "Teachable machine: Approachable Web-based tool for exploring machine learning classification", *InExtended abstracts of the 2020 CHI conference on human factors in computing systems 2020 Apr 25 (pp. 1-8)*.