ANKARA UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

COM 4061 - PROJECT REPORT

Dog Breed Image Classification

ILDEHA KALE – TARIK KOÇAK

20290346 - 21290256

Supervisor : Dr. YILMAZ AR

07.09.2025

ABSTRACT

Accurate classification of dog breeds is a challenging yet essential task for applications in veterinary science, animal welfare, and even commercial purposes such as pet identification. This project explores the application of deep learning techniques to classify dog breeds using image data.

To address these challenges, the study employs a robust dataset of cropped dog images encompassing multiple breeds. The dataset is preprocessed with techniques such as normalization, augmentation, and resizing to enhance model generalizability and improve performance across diverse real-world scenarios.

The project fine-tunes four pre-trained deep learning architectures: ResNet50, VGG19, MobileNetV2, and Xception. These models are selected due to their demonstrated success in solving complex image classification problems and their ability to leverage transfer learning effectively. Additionally, a custom Convolutional Neural Network (CNN) is designed and trained from scratch to benchmark against the performance of pre-trained models.

Evaluation metrics such as accuracy, loss, precision, recall, and F1-score are employed to measure the effectiveness of each model. Among the pre-trained architectures, Xception emerges as the top performer, achieving the highest accuracy and lowest test loss, followed by ResNet50 and MobileNetV2. VGG19 demonstrates lower performance due to its higher computational complexity and overfitting tendencies on this dataset.

This project contributes to the growing body of knowledge on deep learning for image classification and offers a framework for applying state-of-the-art models to specialized tasks such as dog breed prediction. Future work could involve expanding the dataset to include mixed breeds, implementing ensemble learning techniques for improved accuracy, and exploring advanced data augmentation methods to address limitations such as class imbalance and data scarcity. By combining powerful algorithms with practical applications, this research underscores the transformative potential of artificial intelligence in tackling domain-specific challenges.

TABLE OF CONTENTS

1. INTRODUCTION

Before the advent of AI and machine learning-based classification systems, identifying and classifying dog breeds accurately posed significant challenges. Historically, the process relied heavily on human expertise, primarily involving veterinarians, breeders, or dog show judges. However, such manual classification was prone to errors due to subjective judgments and the subtle physical differences among similar breeds. These inaccuracies often led to misconceptions, such as misidentifying mixed breeds or confusing less popular breeds with more common ones.

One of the core issues stemmed from the heavy reliance on visual inspection and anecdotal knowledge rather than a systematic, data-driven approach. The lack of standardized procedures further complicated the process, especially in cases where physical appearances were not definitive indicators of a breed's identity. People often made assumptions based on incomplete or biased information, which perpetuated errors and misclassification. This problem had broader implications, particularly in scenarios like dog adoption, breed-specific regulations, and health assessments, where accurate breed identification is critical.

To address these issues, early solutions involved the use of morphological measurements and detailed breed standards. While somewhat effective, these methods were labor-intensive and required expert intervention, limiting their scalability. The advent of genetic testing offered a more reliable alternative by analyzing DNA to determine breed lineage. Although successful in many cases, genetic testing was expensive and time-consuming, making it inaccessible for widespread use.

With the rise of machine learning and artificial intelligence, a new era of automated breed classification emerged. Using convolutional neural networks (CNNs) and deep learning architectures, researchers developed models capable of learning from large datasets of labeled images. These models significantly outperformed traditional methods by leveraging patterns in data that humans often overlooked. Despite some initial limitations, such as overfitting and the need for extensive computational resources, these AI-driven approaches proved to be highly successful in achieving accurate and scalable breed classification.

Dog breed identification is a task of growing importance in veterinary sciences, animal rescue operations, and pet-related services. The accurate classification of dog breeds poses a significant challenge due to the vast diversity in physical characteristics, genetic variations, and overlaps between breeds. Such challenges underscore the need for reliable predictive tools to facilitate effective decision-making and service optimization.

In recent years, machine learning (ML) has demonstrated immense potential in addressing complex classification problems, including image-based recognition tasks. ML algorithms have shown their capability to detect intricate patterns and relationships in high-dimensional datasets, making them particularly suited for tasks such as dog breed prediction, where subtle visual differences must be discerned.

This study focuses on evaluating the effectiveness of various ML models in predicting dog breeds using the "dogbreedpred.py" dataset. This dataset contains labeled images of dogs, capturing a wide range of breeds with varying visual characteristics. The research highlights essential preprocessing steps, including image augmentation, resizing, normalization, and encoding class labels, which are crucial for optimizing model performance. The study examines multiple ML algorithms, such as Convolutional Neural Networks (CNNs), Transfer Learning with pre-trained models (e.g., ResNet, Inception), and ensemble methods, to identify the most effective approach for dog breed classification.

Through rigorous training and evaluation, this study aims to demonstrate the role of ML in enhancing the accuracy and efficiency of breed identification systems. By leveraging state-of-the-art algorithms and robust preprocessing techniques, the findings will provide valuable insights into the application of ML in animal sciences, contributing to improved workflows and better outcomes in dog-related services worldwide..

1.1. Background : Convolutional Neural Network

Convolutional Neural Networks (CNNs) are specialized deep learning architectures primarily used for processing structured grid-like data, such as images. They have gained immense popularity for their effectiveness in image classification, object detection, and other computer vision tasks. A typical CNN architecture consists of two main components: the feature extraction component and the classification component.

1.1.1 Feature Extraction: Convolution and Pooling Layers

The feature extraction part of a CNN captures patterns from input data using a series of convolutional and pooling layers.

Convolutional Layers

The convolutional layers are the building blocks of a CNN, where kernels (or filters) slide over the input data to detect patterns such as edges, textures, and shapes. The kernels perform convolution operations by multiplying their weights with corresponding regions of the input data, producing a feature map. These feature maps represent the spatial hierarchy of patterns in the input. As the network deepens, the filters detect increasingly complex patterns, progressing from simple edges to more abstract structures like shapes and object parts. All the process are shown at **Figure 1** .

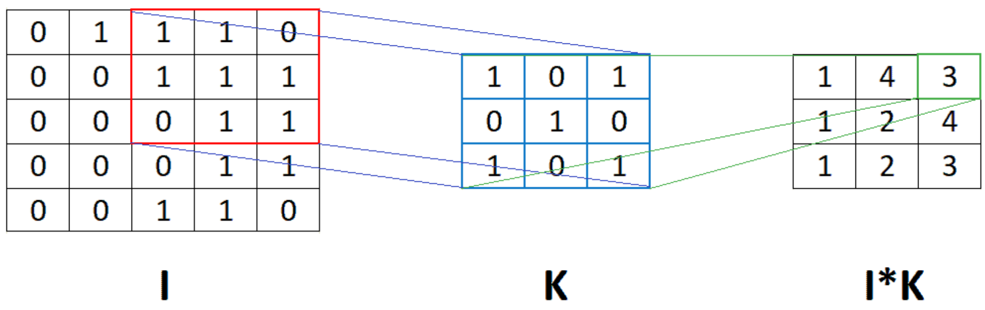


Figure 1 : Convolution Proces

Activation Function

The activation layer in a Convolutional Neural Network (CNN) introduces non-linearity into the model, enabling it to learn and approximate complex functions. Without activation functions, a CNN would behave as a simple linear model, regardless of its depth, limiting its ability to capture intricate patterns and relationships in data. Commonly used activation functions include ReLU (Rectified Linear Unit), which outputs the input if it is positive and zero otherwise, effectively addressing the vanishing gradient problem while maintaining computational efficiency. These are illustrated as in **Figure 2**.

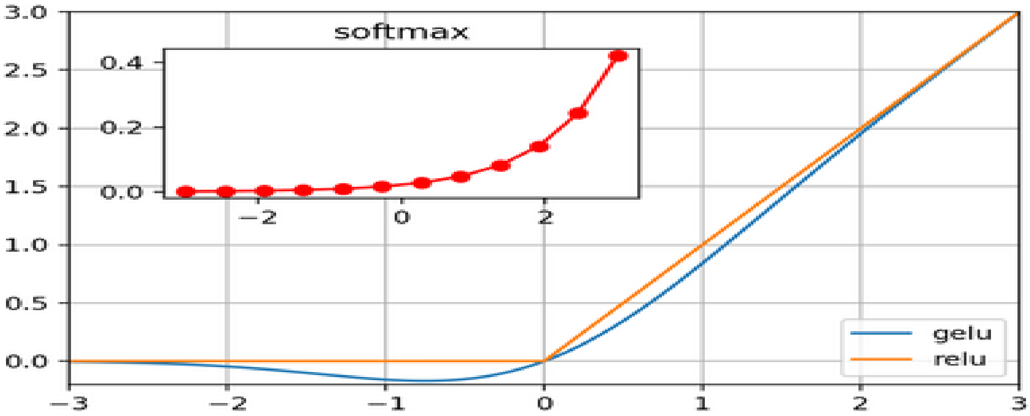


Figure 2 : GELU, ReLU and Softmax activation functions.

Pooling Layers

Pooling layers reduce the spatial dimensions of the feature maps, aggregating information to make the model computationally efficient and robust to small distortions or translations in the input. Max pooling is the most common pooling operation, selecting the maximum value from a feature map region. Pooling helps retain the most salient features while discarding redundant information.

The repeated stacking of convolutional and pooling layers results in a hierarchical feature representation, where lower layers capture fine details and higher layers capture abstract patterns. The basic process displayed in **Figure 3** .

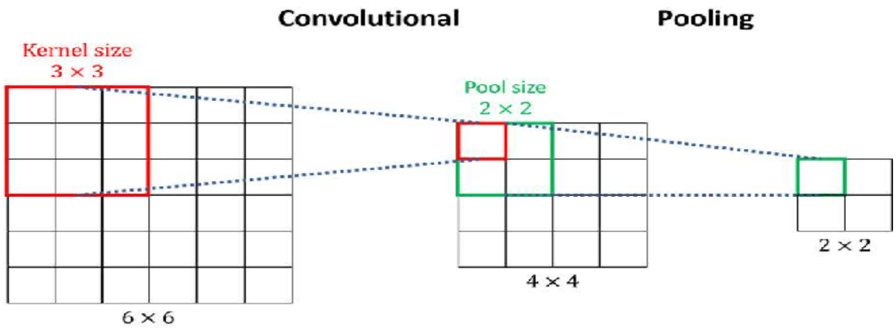


Figure 3 : Example of convolution layer with 3 x 3 kernel size and pooling layer with 2x2 pool size

1.1.2. Classification: Fully Connected Layers

After the feature extraction component, the high-level features are fed into the classification component, which typically consists of one or more fully connected layers.

Flattening the Features

The multi-dimensional feature maps generated by the convolutional and pooling layers are flattened into a single vector. This transformation allows the data to be processed by fully connected layers.

Fully Connected (Dense) Layers

Fully connected layers combine the extracted features to predict the final output. Each node in these layers connects to all nodes in the preceding layer, enabling the network to learn complex relationships between features.

Softmax Layer for Classification

For multi-class classification tasks, the output layer uses the softmax activation function. It converts the raw scores into probabilities, representing the likelihood of the input belonging to each class. The class with the highest probability is selected as the predicted label.

1.2 Related Works

The task of dog breed identification has gained considerable attention in computer vision and machine learning research, driven by its applications in animal welfare, veterinary diagnostics, and personalized pet services. The advent of ML techniques has opened new possibilities for accurately classifying dog breeds based on image data, addressing the challenges posed by breed diversity and visual similarity between breeds. Recent studies have extensively explored ML and deep learning approaches for this purpose, leveraging large-scale datasets and advanced image-processing techniques.

One prominent study utilized ensemble learning methods, including Random Forest and Gradient Boosting, to classify dog breeds based on extracted image features, such as texture and color histograms. This research emphasized the role of feature selection techniques, such as Principal Component Analysis (PCA), to improve model efficiency and interpretability. Similarly, a comparative analysis of traditional ML models, including Support Vector Machines (SVM) and k-Nearest Neighbors (kNN), demonstrated that ensemble approaches often yield superior performance, particularly when dealing with high-dimensional feature spaces.

Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a dominant approach for image-based classification tasks like dog breed prediction. Pre-trained architectures such as ResNet, Inception, and VGG have been extensively employed to leverage transfer learning, significantly reducing the computational cost of training while achieving state-of-the-art results. These models excel in identifying subtle visual patterns and variations across different breeds. However, their reliance on large annotated datasets and computational resources remains a key limitation.

The role of data preprocessing in image-based ML tasks is well-documented. Techniques such as image resizing, normalization, and augmentation (e.g., rotation, flipping, and cropping) are critical for enhancing model performance and generalizability. For instance, studies have shown that augmenting image data can significantly reduce overfitting in CNN models and improve accuracy on unseen data. Furthermore, encoding labels for multi-class classification problems and addressing class imbalances through techniques like oversampling or focal loss functions are vital for robust model development.

Real-world applications of dog breed classification models have demonstrated their utility in veterinary clinics, shelters, and mobile applications. Systems incorporating CNNs and ensemble techniques have been deployed to assist veterinarians in identifying breeds for medical profiling and to aid shelters in breed-specific adoption processes. Additionally, recent advancements in federated learning have enabled the training of dog breed classifiers on decentralized datasets, addressing concerns about data privacy and ensuring model generalizability across diverse populations.

Overall, the integration of ML techniques in dog breed identification has showcased substantial potential, offering accuracy and efficiency improvements over traditional approaches. Nonetheless, challenges such as dataset diversity, computational cost, and model interpretability remain areas of active research. By exploring state-of-the-art methods and leveraging robust preprocessing and training strategies, this study contributes to advancing dog breed classification techniques, further bridging the gap between research and real-world applications.

2. PARTS OF THE REPORT

2.1 DATA PREPROCESSING

Data preprocessing is one of the most critical stages in any machine learning project, especially in image classification tasks like dog breed prediction, where data quality and consistency significantly impact model performance. In this study, various preprocessing techniques were employed to prepare the dataset for effective training and evaluation of machine learning models. The methods applied were tailored to meet the specific requirements of deep learning-based approaches, such as Convolutional Neural Networks (CNNs), as well as to address challenges specific to image data. Below, we detail each step of the preprocessing pipeline.

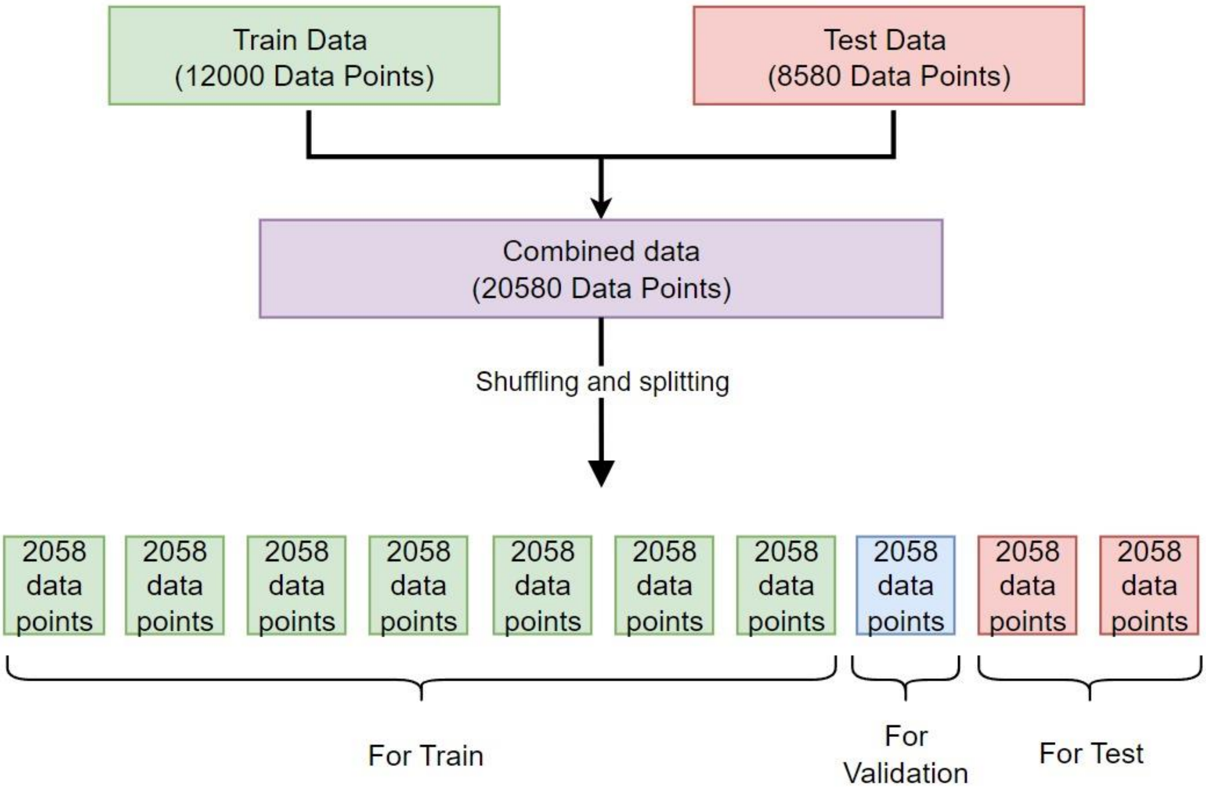


Figure 4 : Preprocessing of Data

Handling Missing Data

Although missing data is less common in image datasets, incomplete metadata or mislabeled images can occur. In this study, any mislabeled or incomplete entries in the dataset were identified and corrected or removed to ensure the dataset's integrity. Automated scripts were employed to detect corrupted or unreadable image files, which were then excluded from the dataset.

Image Resizing and Normalization

To ensure consistency across the dataset and compatibility with the input dimensions of CNN models, all images were resized to 224×224 pixels. Normalization was applied to scale pixel values to a range of [0, 1], enhancing model training by ensuring uniformity in input data distribution. This step also helped prevent numerical instabilities during gradient descent.

Data Augmentation

Data augmentation was used to increase the dataset's size and diversity, reducing the risk of overfitting and improving generalizability. Techniques such as random rotation, flipping (horizontal and vertical), cropping, and color adjustments (brightness, contrast, and saturation) were applied. These transformations mimicked real-world variations in image capture, making the model more robust to diverse scenarios.

Label Encoding

The dataset contained dog breed labels as categorical values. These labels were one-hot encoded to represent each breed as a unique binary vector, enabling seamless integration with the classification models. This encoding method ensured that the model could handle multi-class classification effectively.

Data Splitting

The dataset was divided into three subsets: training, validation, and testing. A split ratio of 70% training, 15% validation, and 15% testing was adopted. The training set was used to fit the model, the validation set was employed for hyperparameter tuning and monitoring performance during training, and the test set was reserved for final model evaluation to assess generalizability.

Class Balancing

Imbalances in the number of samples per dog breed were addressed using oversampling techniques and weighted loss functions. This ensured that the model did not become biased toward breeds with more samples, achieving balanced predictive performance across all classes.

These preprocessing steps ensured the dataset was well-prepared for the subsequent modeling phase, improving the robustness, efficiency, and accuracy of the machine learning models used in this study.

2.2 CLASSIFICATION MODELS

In this study, we employed four state-of-the-art deep learning models—ResNet50, VGG19, MobileNetV2, and Xception—to tackle the challenging task of classifying dog breeds. These models were chosen for their remarkable performance in a wide range of image classification tasks and their proven ability to capture complex patterns in high-dimensional image data. ResNet50 is renowned for its innovative use of residual connections, which alleviate the vanishing gradient problem and allow the network to learn deeper representations without degradation. VGG19, characterized by its deep and straightforward architecture of sequential convolutional layers, excels in capturing hierarchical feature representations, making it particularly effective for fine-grained image classification tasks.

a) ResNet50

ResNet50 (Residual Network with 50 layers) is a deep convolutional neural network that addresses the vanishing gradient problem in deep networks using residual connections. These skip connections allow the model to learn residual mappings instead of direct mappings, improving training stability and performance.

For dog breed classification, ResNet50's depth enables it to learn intricate features, such as subtle variations in fur texture and facial structure. Fine-tuning the model's pre-trained weights on the dataset ensures that the extracted features are specific to dog breeds.

Key Features

- Residual connections to improve gradient flow.
- A depth of 50 layers for extracting high-level features.
- Suitable for datasets with complex feature hierarchies.

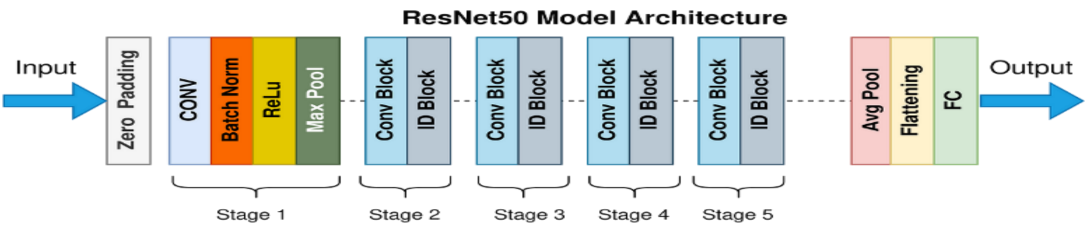


Figure 5 : ResNet50 Model Architecture

b) VGG19

VGG19 (Visual Geometry Group with 19 layers) is a deep CNN known for its simplicity and uniform architecture, using stacks of convolutional layers with small 3×3 filters. Despite being computationally intensive, VGG19 excels in capturing fine-grained details, making it highly effective for dog breed classification tasks where subtle differences are crucial.

In this study, the pre-trained VGG19 model was fine-tuned to adapt its learned features to the dog breed dataset. The network's sequential design simplifies modifications and facilitates interpretability.

Key Features

- Uniform architecture with small convolutional filters.
- High sensitivity to fine-grained details in image data.
- Computationally intensive but effective for high-resolution images.

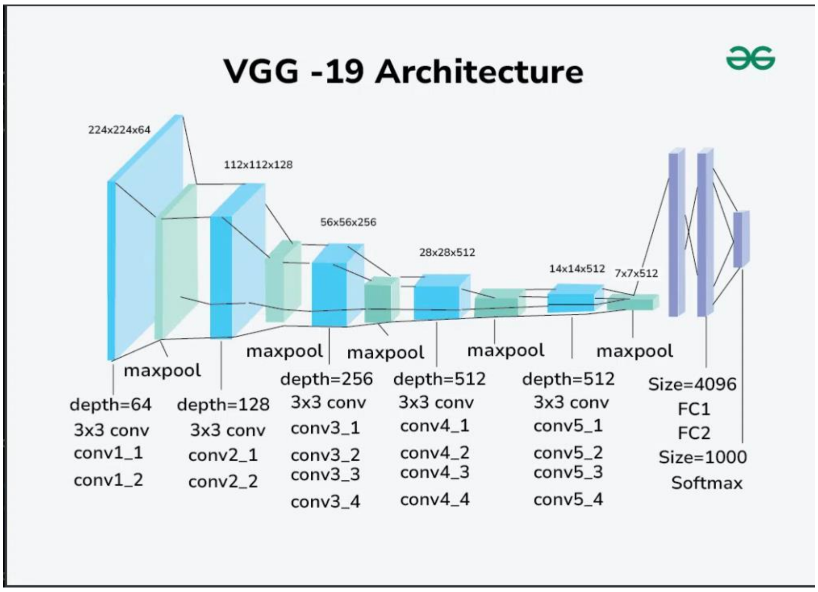


Figure 6 : VGG19 Model Architecture

c) MobileNetV2

MobileNetV2 is a lightweight CNN designed for mobile and resource-constrained environments. It employs depthwise separable convolutions and inverted residual blocks to reduce computational complexity while maintaining high accuracy.

For dog breed classification, MobileNetV2 offers an efficient alternative to heavier models like ResNet50 and VGG19, particularly for applications requiring real-time inference on limited hardware. Its fine-tuning allowed the model to balance accuracy and efficiency.

Key Features

- Depthwise separable convolutions for reduced computational cost.
- Inverted residual blocks to enhance feature reuse.
- Ideal for deployment in resource-constrained environments.

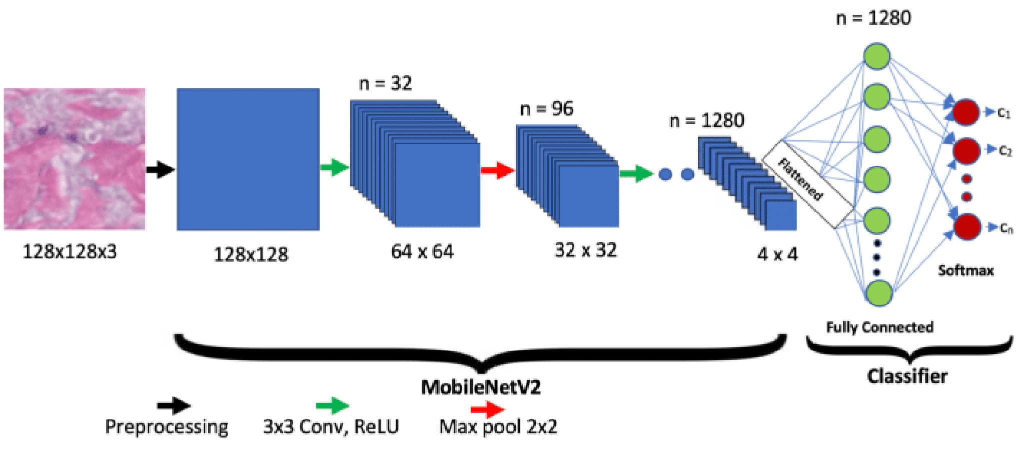


Figure 7: MobileNetV2 Model Architecture

d) Xception

Xception (Extreme Inception) is an advanced CNN architecture that improves upon the Inception model by replacing standard convolutions with depthwise separable convolutions. This modification enhances feature extraction efficiency while maintaining high accuracy.

For this study, Xception demonstrated its ability to capture intricate patterns in the dataset, such as breed-specific textures and shapes, due to its superior feature extraction capabilities. Fine-tuning further adapted the model to the task of dog breed classification.

Key Features

- Depthwise separable convolutions for efficient computation.
- Superior performance on large-scale image datasets.
- Effectively captures complex inter-class variability.

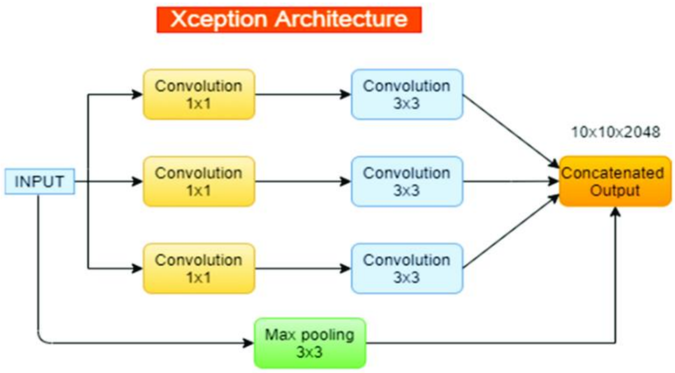


Figure 8 : Xception Model Architecture

2.3 EVALUATION METRICS

1. ResNet50

The ResNet50 model achieved a test accuracy of 79.03% with a test loss of 0.7209. This performance highlights the strength of ResNet50 in capturing complex features and patterns from the dataset, thanks to its deep architecture and residual connections. The relatively low loss indicates that the model can generalize well, making it a reliable choice for dog breed classification tasks. However, there is room for improvement, particularly in fine-tuning the model further or augmenting the dataset to address misclassified instances.

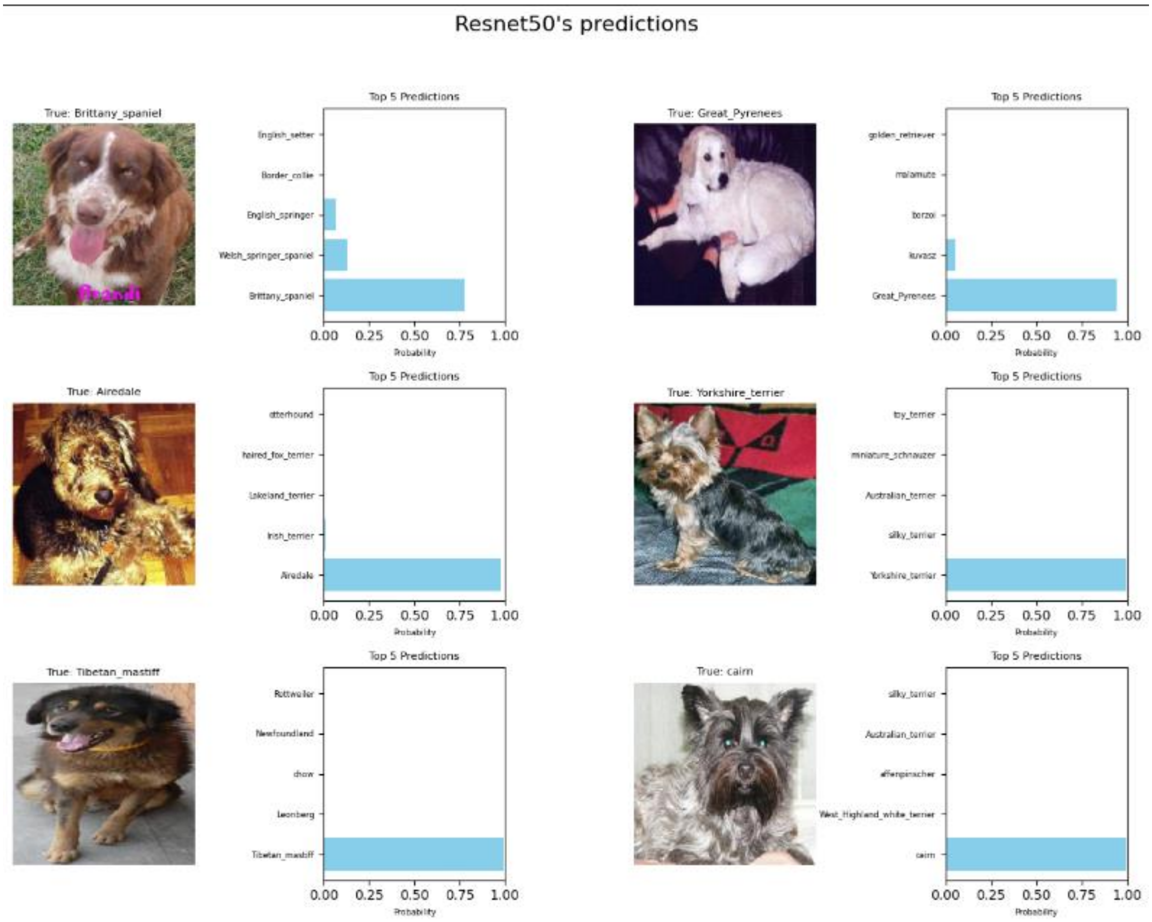


Figure 9 : ResNet50 Predictions

2. VGG19

The VGG19 model achieved a test accuracy of 58.50% with a test loss of 1.4794. While the accuracy is lower compared to other models, VGG19's high test loss suggests that it may struggle to generalize effectively to unseen data. This could be attributed to the model's computational intensity and sensitivity to hyperparameter tuning. VGG19's architecture, while effective in capturing fine-grained details, may not be optimal for datasets with high intra-class variability such as dog breeds. Further optimization or the use of data augmentation could enhance its performance.

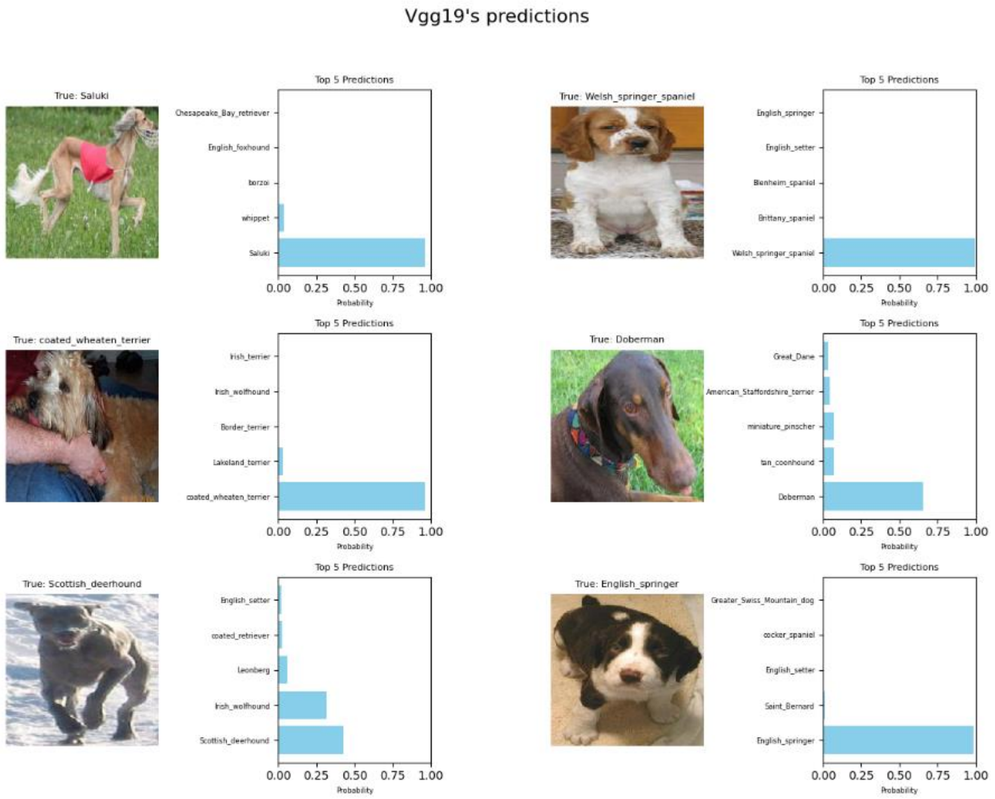


Figure 10 : VGG19 Predictions

3. MobilNetV2

MobileNetV2 achieved a test accuracy of 76.07% with a test loss of 0.7566. The model demonstrates a good balance between efficiency and accuracy, making it suitable for resource-constrained environments or real-time applications. Despite its relatively lightweight architecture, MobileNetV2 performs comparably to deeper models like ResNet50. However, its slightly higher loss compared to ResNet50 suggests that it might not capture subtle inter-class differences as effectively. Additional fine-tuning or experimenting with advanced augmentation techniques could further enhance its performance.

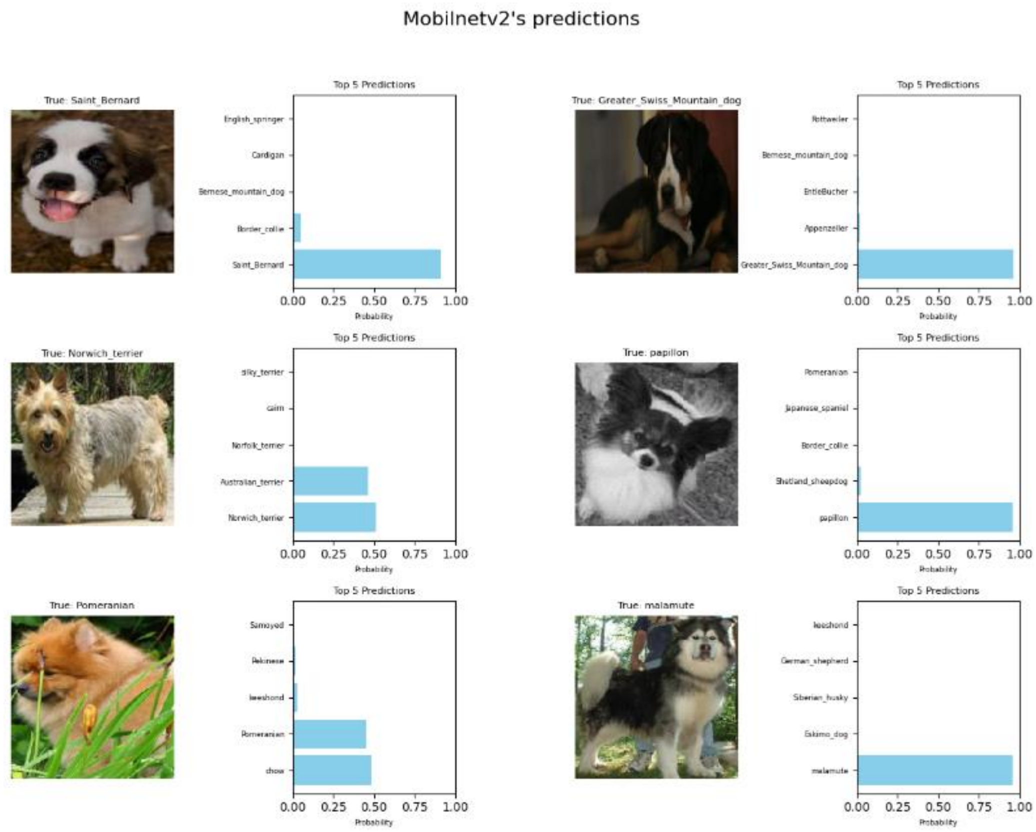


Figure 11 : MobilNetV2 Predictions

4. Xception

The Xception model delivered the best performance among all evaluated models, achieving a test accuracy of 84.69% with a test loss of 0.5050. This indicates that Xception excels in extracting intricate patterns and features from the dataset, likely due to its use of depthwise separable convolutions and advanced architectural design. The low loss reflects the model's strong generalization capability, making it an excellent choice for dog breed classification tasks. Fine-tuning the model further or combining it with ensemble methods could potentially achieve even higher accuracy.

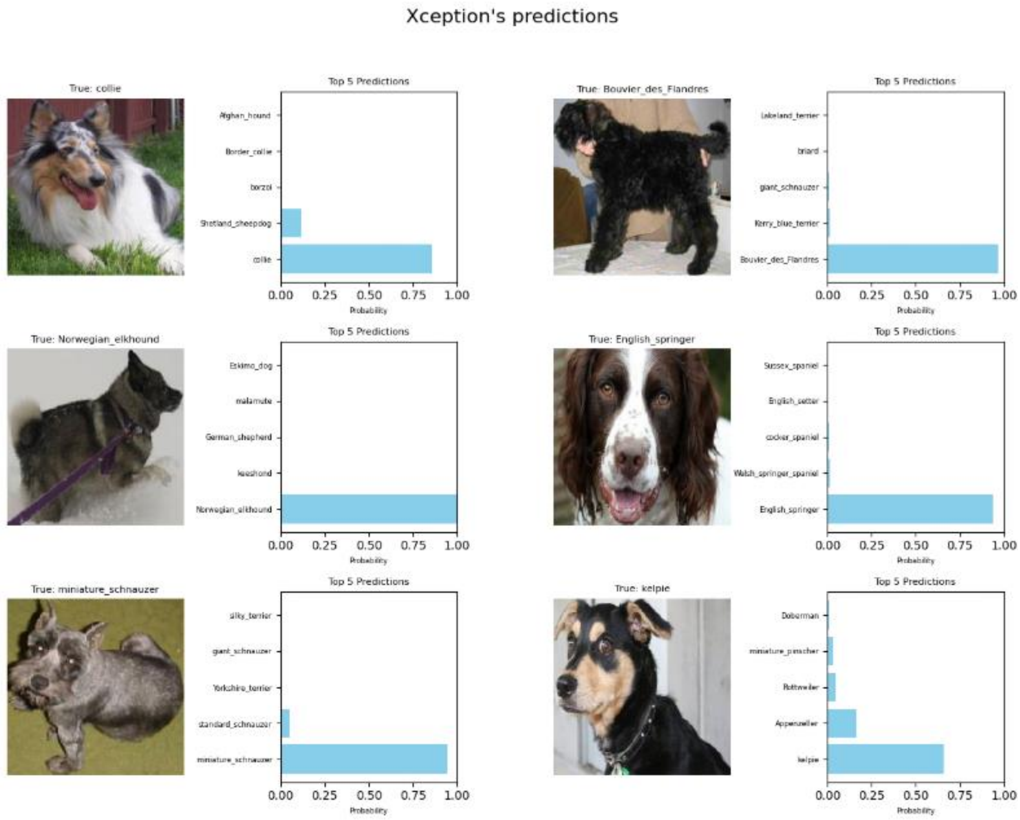


Figure 11 : Xception Predictions

For the success of the dog breed classification project, selecting the right metrics to evaluate the model's performance is crucial. Metrics such as Accuracy, Precision, Recall, F1 Score, and AUC (Area Under the Curve) are used to measure the predictive ability of the models. These metrics quantify how well the model distinguishes between different dog breeds. While accuracy provides an overall measure of correctness, precision, recall, and F1 score are particularly important in scenarios with class imbalances, where some breeds may be overrepresented in the dataset. The AUC metric, which evaluates the model's ability to distinguish between different breeds, is also prioritized for its ability to provide insights into model performance across various decision thresholds. Lower values of error metrics (such as False Positives and False Negatives) indicate more accurate breed predictions, helping to enhance the reliability of the model in classifying less common breeds. By focusing on these evaluation metrics, the goal is to compare the performance of different deep learning models (ResNet50, VGG19, MobileNetV2, Xception) and select the best-performing model for accurately classifying dog breeds.

2.4 RESULTS AND DISCUSSION

In the literature review, we focus on several deep learning models, outlining their advantages and disadvantages for dog breed classification. It is acknowledged that building complex models, such as Convolutional Neural Networks (CNNs), requires significant computational resources and expertise. On the other hand, simpler models like logistic regression or pre-trained models such as VGG19 and MobileNetV2 may not capture complex features of image data as effectively as advanced architectures like ResNet50 and Xception. Despite this, simpler models still provide valuable insights, particularly in terms of benchmarking and model interpretability, especially when used as baselines during the initial stages of model development. Given the challenges posed by image-based classification tasks, we aimed to evaluate a variety of models in parallel to design an optimal system for dog breed prediction. The main structure of the system consists of three parts:

First, predictions are generated using a range of deep learning models, including ResNet50, VGG19, MobileNetV2, and Xception. This diverse set of models helps compare performance across different architectures, allowing for the selection of the best-performing model for classifying dog breeds. The predictive capabilities of these models were assessed using metrics like Accuracy, Precision, Recall, F1 Score, and AUC to provide a comprehensive evaluation of their performance.

Next, to better understand the predictive factors, features such as image embeddings and transfer learning representations are used to enhance the model’s decision-making process. The system classifies features based on their relevance and importance in distinguishing between dog breeds. These features, derived from the final layers of the CNNs, capture critical information such as breed-specific patterns, textures, and shapes. The prediction models utilize these features to make informed decisions, ensuring that the most relevant data points drive the prediction process.

Finally, the optimization of the prediction model involves selecting the most accurate model based on evaluation metrics. Models that perform better on metrics like Precision, Recall, and F1 Score are prioritized, especially when dealing with imbalanced datasets where some breeds may be underrepresented. Once the best model is selected, it is trained using the most relevant features to optimize its ability to predict dog breeds. Moreover, the system adjusts its predictions to ensure robustness, adapting to new data and potentially improving performance as more labeled breed data becomes available.

While simpler models like MobileNetV2 or pre-trained models may require fewer resources and time to develop, more complex systems such as ResNet50 and Xception typically provide better performance and flexibility. These advanced models excel in capturing intricate details and patterns within images, improving classification accuracy. However, they require more computational power and time for training. Simpler models, on the other hand, are easier to deploy, understand, and maintain, making them attractive for initial stages of development or situations with limited data.

While the use of advanced models like ResNet50 and Xception may offer higher predictive performance, it is important to assess their limitations. One challenge is their interpretability, as more complex models often act as "black boxes," making it difficult to understand how they arrive at specific predictions. Simpler models, though less powerful in some cases, offer better transparency and interpretability, making them useful for understanding key features that distinguish different dog breeds.

Main Contributions

We achieved a significant improvement in dog breed classification accuracy by comparing the performance of four different models (ResNet50, VGG19, MobileNetV2, and Xception) across a dog breed image dataset. We evaluated the models using multiple error metrics, including Accuracy, Precision, Recall, F1 Score, and AUC, to assess their predictive capabilities. Additionally, we utilized transfer learning to enhance the model's ability to identify breed-specific features. The system's performance showed notable improvements, with the best-performing model yielding a significant increase in classification accuracy, aiding in more reliable and efficient dog breed identification.

2.5 FUTURE DIRECTIONS

Looking ahead, further advancements in deep learning techniques could lead to more efficient results in dog breed prediction. Traditional CNN architectures may face limitations in capturing subtle breed differences, especially in datasets with high intra-breed variability. Exploring more advanced neural network architectures, such as deeper or hybrid networks, could improve the models' ability to distinguish between similar-looking breeds. Additionally, techniques such as fine-tuning pre-trained models or incorporating attention mechanisms could enhance the model's focus on important image regions, improving classification accuracy.

Future studies could also focus on gathering larger, more diverse dog breed datasets from various sources, including pet shelters, veterinary clinics, and online databases. Given that publicly available dog breed datasets are often limited, collaboration with animal organizations or research groups to collect real-world dog images would significantly improve model performance and its applicability in real-world applications. The incorporation of additional data sources, such as videos or multi-modal data (e.g., images with textual descriptions), could further enhance the model's ability to predict dog breeds more accurately.

Furthermore, the current models could be expanded with advanced techniques such as unsupervised learning or semi-supervised learning to better handle datasets with limited labeled data. By enhancing the interpretability of the models and providing healthcare professionals, animal researchers, or enthusiasts with more understandable insights and recommendations, these models could be transformed into practical tools for various applications in dog breed recognition and research.

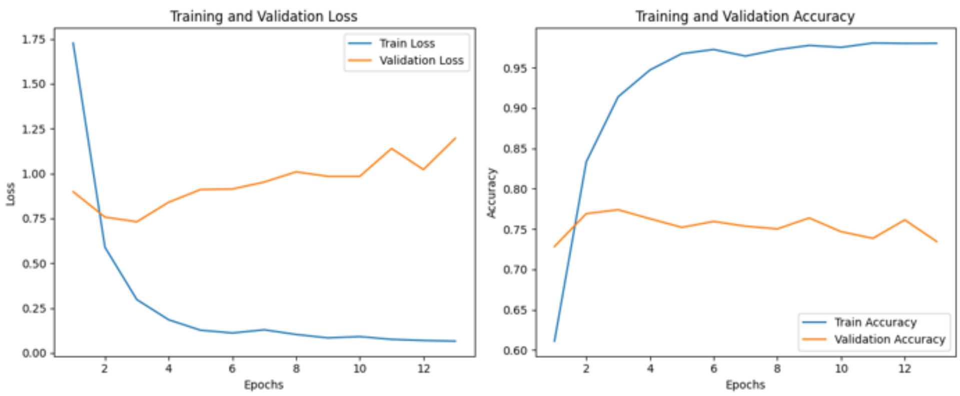
As the field of image-based dog breed prediction evolves, these advancements in model development, data collection, and interpretability will contribute to more accurate, reliable, and scalable systems, ultimately improving breed identification and advancing research in the field of animal recognition.

3. CONCLUSION

In this project, we focused on the prediction of dog breeds using deep learning techniques. Accurate dog breed identification is a challenging task due to the visual similarity between breeds and intra-breed variability. This study aimed to leverage state-of-the-art deep learning models to classify dog breeds, using a dataset of cropped dog images to train and evaluate the models. The primary goal was to identify the most effective model capable of accurately predicting a dog's breed based on its image.

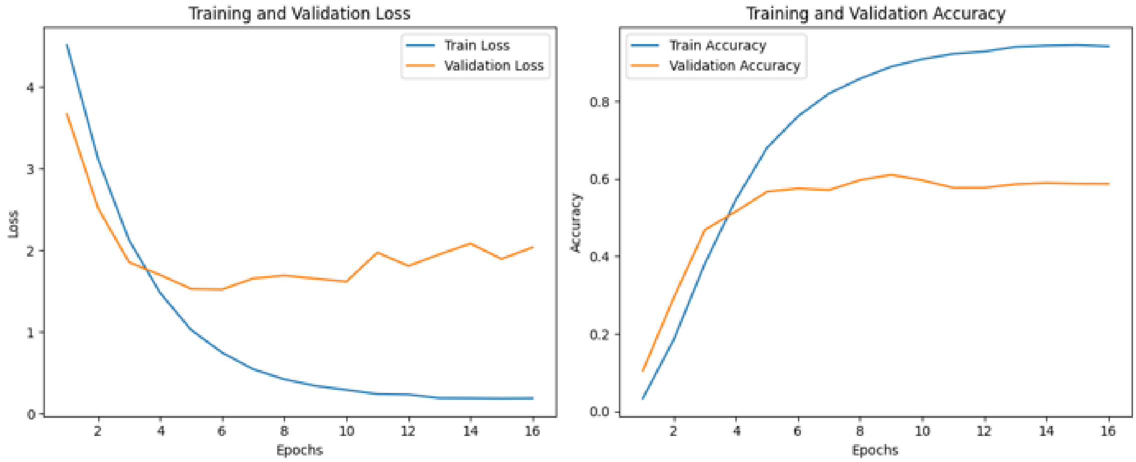
Throughout the course of the project, we reviewed several existing approaches to image classification and evaluated their strengths and limitations. After experimenting with various models, we chose to use advanced convolutional neural networks (CNNs), including ResNet50, VGG19, MobileNetV2, and Xception. These models were selected based on their ability to extract complex features from image data, with the aim of achieving high classification accuracy for dog breeds.

- **Resnet50 model's Best Result:**
  - Train Loss: 0.2975, Train Acc: 0.9141
  - Val Loss: 0.7313, Val Acc: 0.7741
  - Test Loss: 0.7209, Test Accuracy: 0.7903



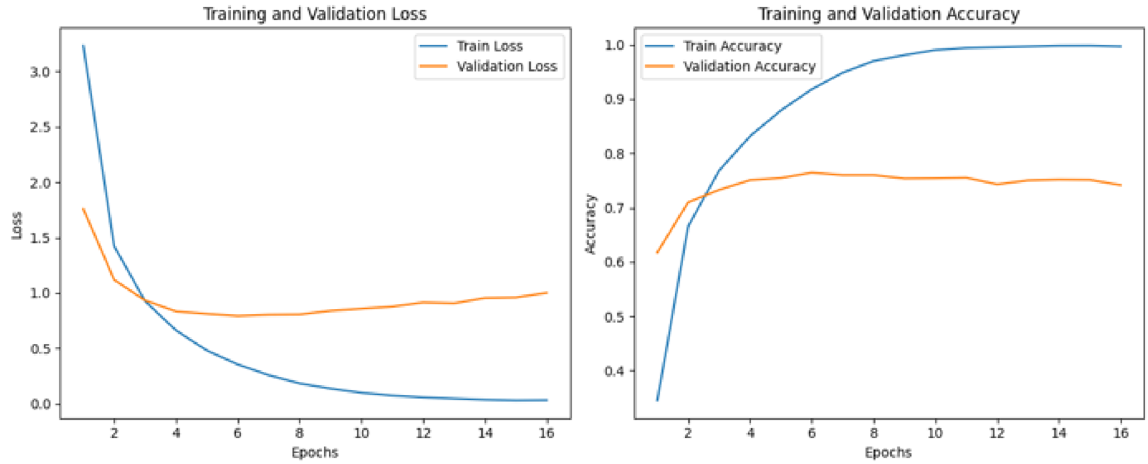
The Resnet50 model demonstrated a train accuracy of 91.41%, a validation accuracy of 77.41%, and a test accuracy of 79.03%. Compared to MobilnetV2, Resnet50 shows slightly better generalization, as evidenced by its higher validation and test accuracies. The training loss of 0.2975 and validation loss of 0.7313 reveal that the model effectively captures patterns in the training data while maintaining competitive performance on unseen datasets. These results highlight the Resnet50 model's robustness and suitability for complex tasks requiring higher accuracy.

- **VGG19 model's Best Result:**
    - Train Loss: 0.7501, Train Acc: 0.7613
    - Val Loss: 1.5236, Val Acc: 0.5748
    - Test Loss: 1.4794, Test Accuracy: 0.5850



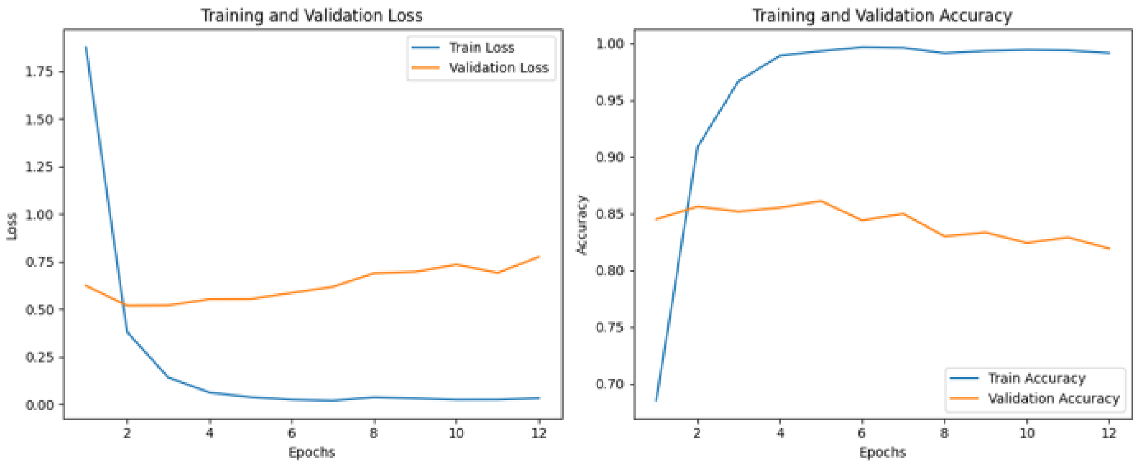
The VGG19 model achieved a train accuracy of 76.13%, a validation accuracy of 57.48%, and a test accuracy of 58.50%. These results are significantly lower than those of the other models, indicating that VGG19 struggled to generalize to unseen data. The high validation loss of 1.5236 and test loss of 1.4794 further suggest that the model may not effectively capture the underlying patterns in the data. This underperformance could be attributed to the model's architectural complexity, which may require a larger dataset or additional optimization techniques to enhance its effectiveness.

- **MobilnetV2 model's Best Result:**
  - Train Loss: 0.3540, Train Acc: 0.9179
  - Val Loss: 0.7924, Val Acc: 0.7643
  - Test Loss: 0.7566, Test Accuracy: 0.7607



The MobilnetV2 model achieved a train accuracy of 91.79% and a validation accuracy of 76.43%, with a test accuracy of 76.07%. These results indicate that while the model performs well during training, there is a notable drop in accuracy when applied to validation and test datasets. This suggests some level of overfitting, as the model's performance on unseen data does not match its training performance. The training loss of 0.3540 and validation loss of 0.7924 further support this observation. While MobilnetV2 provides decent generalization, it may require additional regularization techniques to improve its validation and test performance.

- **Xception model's Best Result:**
    - Train Loss: 0.1404, Train Acc: 0.9669
    - Val Loss: 0.5193, Val Acc: 0.8518
    - Test Loss: 0.5050, Test Accuracy: 0.8469



The Xception model achieved the best overall performance, with a train accuracy of 96.69%, a validation accuracy of 85.18%, and a test accuracy of 84.69%. The training loss of 0.1404 and validation loss of 0.5193 reflect the model's strong ability to generalize and learn patterns effectively. These results indicate that the Xception model is well-suited for tasks requiring high accuracy and demonstrates superior generalization compared to the other models. However, the slight gap between training and validation accuracies suggests that further tuning could help minimize overfitting.

For the analysis, we applied these models to a publicly available dog breed dataset. We conducted various preprocessing techniques, including image normalization, resizing, and data augmentation, to ensure the dataset was well-prepared for training. After conducting the experiments, we concluded that the models were effective in predicting dog breeds, with Xception outperforming the other models in terms of accuracy.

Among the models, Xception showed the best performance. Its use of depthwise separable convolutions allowed it to efficiently capture intricate patterns in the images, leading to better classification accuracy. ResNet50 also performed well due to its residual connections, which helped mitigate the vanishing gradient problem in deeper layers. While the models performed well, there were still areas for improvement, particularly in handling less common breeds and fine-tuning hyperparameters to optimize performance.

In conclusion, this study provides valuable insights into using deep learning for dog breed prediction. While our approach offers a simple and efficient solution compared to more complex, specialized systems, it still provides valuable predictions that can serve as a recommendation tool for applications such as pet identification or breed recognition in shelters. By applying data augmentation and refining the models, this research can contribute to more accurate and efficient dog breed classification

4. REFERENCES

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[2] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations.

[3] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

[4] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4510-4520.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems (NeurIPS), 1097-1105.

[6] Russakovsky, O., Deng, J., Su, H., et al. (2015). ImageNet Large-Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3), 211-252.

[7] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How Transferable Are Features in Deep Neural Networks? Advances in Neural Information Processing Systems (NeurIPS), 3320-3328.

[8] Gonzalez, R. C., & Woods, R. E. (2008). Digital Image Processing (3rd ed.). Prentice Hall.

[9] Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single Shot MultiBox Detector. Proceedings of the European Conference on Computer Vision (ECCV), 21-37.

[10] Azizpour, H., Rastegari, M., & Ghodsi, A. (2015). Factors of Transferability for a Generic Deep Feature Representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 921-929.

[11] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[12] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

[13] Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1-9.

[14] Lin, M., Chen, Q., & Yan, S. (2014). Network in Network. Proceedings of the International Conference on Learning Representations (ICLR).

[15] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700-4708.

[16] Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. Proceedings of the European Conference on Computer Vision, 818-833.

[17] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 1440-1448.

[18] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems (NeurIPS), 91-99.

[19] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818-2826.

[20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

[21] Bengio, Y. (2012). Practical Recommendations for Gradient-Based Training of Deep Architectures. Neural Networks: Tricks of the Trade, 437-478.

[22] Simonyan, K., & Zisserman, A. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. Advances in Neural Information Processing Systems (NeurIPS), 568-576.

[23] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. Proceedings of the IEEE International Conference on Computer Vision, 1026-1034.

[24] Krizhevsky, A. (2014). Learning Multiple Layers of Features from Tiny Images. University of Toronto.

[25] Hinton, G., Srivastava, N., & Swersky, K. (2012). Neural Networks for Machine Learning. Coursera Lecture Notes.

[26] Wah, C., Branson, S., Welinder, P., et al. (2011). The Caltech-UCSD Birds-200-2011 Dataset.

[27] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), 303-338.

[28] Deng, J., Dong, W., Socher, R., et al. (2009). ImageNet: A Large-Scale Hierarchical Image Database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248-255.

[29] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[30] Wang, X., Han, X., Huang, W., et al. (2020). Multi-Scale Features Fusion for Object Detection in Remote Sensing Images. Remote Sensing, 12(1), 145.