AI Generated Art vs Real Art

Objective:

The "AI Generated Art vs Real Art" dataset is curated to advance the understanding, detection, and creative utilization of synthetic and authentic visual content. By integrating a diverse collection of web-scraped real-world images with AI-generated visuals produced by modern generative models, the dataset offers a comprehensive and balanced foundation for machine learning research, digital forensics, and artistic experimentation.

This dataset spans a wide range of visual domains, including people, animals, landscapes, portraits, psychedelic visuals, and paintings, enabling robust model development, generalization, and creative synthesis.

Unified Core Objectives :

1. **Binary Image Classification (AI vs Real):**
   Facilitate the development, training, and evaluation of machine learning and deep learning models to accurately classify images as either AI-generated or real, promoting advancements in AI transparency, digital content verification, and synthetic media detection.
2. **Cross-Domain Generalization:**
   Provide a visually diverse dataset to ensure models can generalize effectively across a variety of image categories and styles, supporting real-world applicability and adaptability.
3. **Visual Pattern Analysis and Feature Discovery:**
   Support data scientists and researchers in conducting exploratory data analysis (EDA) to uncover trends, stylistic differences, texture artifacts, and feature correlations that distinguish synthetic content from authentic images.
4. **Creative and Artistic Exploration:**
   Empower artists, designers, and media practitioners to utilize the dataset as a springboard for computational aesthetics, blending traditional techniques with AI-generated visuals to produce innovative artworks and multimedia experiences.
5. **AI Literacy, Education, and Digital Forensics:**
   Contribute to the broader field of AI ethics and media literacy by supporting the creation of educational tools and forensic technologies capable of identifying manipulated or machine-generated content — critical for combating misinformation and understanding emerging visual AI systems.

Introduction:

The rapid advancement of artificial intelligence (AI) has led to the creation of synthetic images that are increasingly indistinguishable from real photographs. Tools like DALL·E, MidJourney, and Stable Diffusion have enabled the generation of hyper-realistic visuals, raising critical concerns about authenticity, trust, and the implications of synthetic media. In response to these challenges, the "AI Generated Images vs Real Images" dataset has been curated to provide a comprehensive resource for distinguishing between authentic and machine-generated visual content.

This dataset comprises a diverse collection of images sourced from two distinct channels: real-world web-scraped imagery and AI-generated content. The images span a wide array of subjects, including people, animals, portraits, landscapes, and psychedelic visuals, offering a rich and varied dataset for both technical analysis and artistic exploration.

By providing access to this dataset, the goal is to facilitate the development and evaluation of machine learning models capable of accurately classifying images as real or AI-generated. Additionally, it serves as a foundation for exploratory data analysis, enabling researchers to uncover visual patterns and features that differentiate authentic images from synthetic ones. Furthermore, the dataset supports creative and artistic experimentation, allowing artists and designers to explore new avenues of expression by blending AI-generated imagery with real-world visuals.

In this study, we utilize the "AI Generated Images vs Real Images" dataset to investigate classification performance across various model architectures. We also analyze visual characteristics that help distinguish real imagery from machine-generated visuals. As part of our broader contribution to the field, we aim to benchmark baseline performance, propose potential augmentation techniques, and suggest practical applications of the dataset in domains such as media integrity and AI safety.

Related Works:

The rapid advancements in artificial intelligence, particularly in generative models for image synthesis, have led to an increasing concern regarding the authenticity of visual content. This section reviews relevant research on AI-generated images, deepfake detection, and the role of datasets in advancing the detection of synthetic media.

1. AI-Generated Images and Generative Models

Recent breakthroughs in AI-based image generation, particularly through models such as Generative Adversarial Networks (GANs) [1], Variational Autoencoders (VAEs) [2], and Diffusion Models like Stable Diffusion [3] and DALL·E [4], have raised new challenges in distinguishing authentic images from those generated by machines. These models are capable of producing highly realistic images across various domains, ranging from portraits and landscapes to abstract and psychedelic visuals. The ability of these models to generate content with visual realism has made it increasingly difficult for both humans and machines to identify whether an image is AI-generated or real.

The works of Radford et al. (2021) with CLIP [5] and Rombach et al. (2022) on Latent Diffusion Models [6] have demonstrated significant strides in generating images that often surpass human-generated content in terms of creativity and visual appeal, pushing the boundaries of what can be considered "authentic" in digital media.

2. Detection of AI-Generated and Deepfake Images

As synthetic media becomes more pervasive, the need for accurate detection tools has intensified. Early approaches to deepfake detection primarily focused on identifying manipulated videos, but recent research has expanded to include image-based methods for detecting AI-generated images. For instance, Matern et al. (2020) proposed a deep learning model capable of detecting fake images generated by GANs by identifying inconsistencies in textures, lighting, and other features [7]. Similarly, Zhou et al. (2021) introduced a method using convolutional neural networks (CNNs) for detecting deepfake images by analyzing pixel-level artifacts and inconsistencies in generated faces [8].

In the context of AI-generated images from models like DALL·E and MidJourney, recent work by Shao et al. (2022) [9] focused on training CNNs to recognize visual cues such as texture, lighting, and pixel artifacts that distinguish AI-generated images from real ones. These models leverage large datasets that contain both synthetic and real images to learn the subtle differences between the two, further advancing the field of synthetic media detection.

3. Datasets for Fake Image Detection

Datasets are a critical resource for training and evaluating AI models in the domain of synthetic media detection. One of the first prominent datasets for fake image detection was DeepFake Detection Challenge (DFDC) dataset [10], which focuses on deepfake video content but has inspired similar efforts in the image domain. In a parallel vein, the FF++ (FaceForensics++)

dataset [11] has been instrumental in developing algorithms capable of detecting facial manipulations in both video and still images.

A more recent and directly relevant dataset is the Fake Image Detection Dataset introduced by Li et al. (2023), which contains a balanced collection of AI-generated and real images specifically designed to facilitate binary classification between real and fake images [12]. This dataset offers a broad variety of image types, including portraits, landscapes, and animals, which aligns closely with the AI Generated Images vs Real Images dataset in terms of its diversity and scope.

4. Challenges and Advancements in Image Classification

Despite significant advancements, detecting AI-generated images remains a challenging task. The high-quality output produced by generative models, combined with the variety of techniques used to generate them, introduces significant complexity in detection. Models such as GANs can produce images with nearly flawless textures and facial features, often making it difficult for traditional methods to distinguish between real and synthetic images. Additionally, the diversity of generation methods, including StyleGAN2 [13], BigGAN [14], and VQ-VAE-2 [15], further complicates detection, as each model introduces unique artifacts that require specialized models for classification.

Recent research has introduced hybrid approaches, such as multimodal fusion and ensemble learning [16], which combine multiple detection techniques to improve accuracy and generalization. For instance, models using transfer learning [17] have shown great promise in adapting pre-trained networks to the task of fake image detection, leading to improved detection performance across a variety of domains.

5. Contribution of the AI Generated Images vs Real Images Dataset

The AI Generated Images vs Real Images dataset stands as a crucial step in addressing the growing need for high-quality resources to train AI models for synthetic media detection. Unlike traditional datasets that are either biased toward one type of generative model or limited to a specific category of images, this dataset offers a comprehensive mix of real and AI-generated content across various domains. This diversity enhances the dataset's utility in developing models capable of generalizing across different visual domains, such as portraits, landscapes, animals, and psychedelic art, ensuring broader applicability in real-world scenarios.

Additionally, this dataset supports creative and exploratory applications, enabling artists and designers to engage with AI-generated content in novel ways, while also providing the research community with a robust foundation for developing more accurate and adaptable detection models.

Limitations:

While the AI Generated Images vs Real Images dataset provides a valuable resource for advancing the detection of synthetic media, there are several limitations that should be considered when using it for research and development purposes:

1. **Limited Variety of AI Models:** The dataset includes AI-generated images from a selection of popular generative models, such as Stable Diffusion, DALL·E, and others. However, it does not encompass the full spectrum of generative models available today, such as BigGAN, StyleGAN2, or VQ-VAE-2, which produce distinct types of artifacts and features. As a result, models trained on this dataset may not generalize well to images generated by less common or emerging AI models.

2. **Domain-Specific Biases:** While the dataset spans multiple categories, including portraits, landscapes, animals, and psychedelic art, it may still exhibit biases toward certain visual domains or artistic styles. For example, the dataset might underrepresent certain visual genres or specific use cases, such as fashion, architecture, or highly realistic medical images, which could limit its applicability to broader or more niche detection tasks.

3. **Image Quality Variations:** The dataset includes both high-quality and lower-resolution images, reflecting the varying performance of generative models. In some cases, the AI-generated images may exhibit noticeable artifacts or defects that may be easier to detect. However, as generative models continue to improve, the gap between synthetic and real images will likely shrink, making it challenging to maintain a consistent level of difficulty across all images in the dataset.

4. **Ethical and Privacy Concerns:** While the dataset is carefully curated, there may be ethical and privacy considerations regarding the web-scraped real images. In some cases, the dataset could contain images of people that were scraped without consent, raising privacy issues for individuals featured in the content. This limitation highlights the importance of ensuring proper ethical guidelines in future datasets.

5. **Lack of Temporal Dynamics:** The dataset primarily consists of static images, without a focus on temporal consistency or the evolution of synthetic media over time. As generative models improve, they may produce more dynamic or contextually complex

images, such as video or interactive media. This limitation restricts the dataset's utility for research in dynamic media verification.

Contributions:

Despite these limitations, the AI Generated Images vs Real Images dataset offers significant contributions to the field of synthetic media detection and AI research:

1. **Comprehensive Dataset for AI vs Real Image Classification:** The dataset serves as a valuable resource for training and evaluating machine learning models focused on the binary classification task of distinguishing AI-generated images from real ones. By offering a diverse mix of image categories and generative models, it enables researchers to develop more generalizable models that can adapt to various types of images and contexts.

2. **Facilitating AI Transparency and Accountability:** As generative models become increasingly sophisticated, the need for transparency in AI-generated content grows. This dataset supports the development of tools and algorithms that can identify AI-generated media, contributing to efforts to ensure accountability and authenticity in digital content.

3. **Enhancing AI Literacy and Digital Forensics:** By providing a curated collection of AI-generated and real images, the dataset helps promote AI literacy and understanding of synthetic media among a broader audience, including media professionals, educators, and policy makers. It also plays a crucial role in the field of digital forensics, where the ability to distinguish between real and fake media is essential for ensuring the integrity of digital evidence and preventing the spread of misinformation.

4. **Supporting Cross-Domain Research:** The diversity of the dataset, which includes a wide range of visual domains such as portraits, animals, and abstract psychedelic imagery, supports cross-domain research in image recognition, feature analysis, and style-based classification. This encourages the development of more robust and flexible models capable of handling different types of visual content.

5. **Contributing to Artistic Exploration:** Beyond technical applications, the dataset serves as a valuable resource for artists, designers, and creators exploring the intersection of AI

and human creativity. By blending real and synthetic images, it offers unique opportunities for artistic exploration, digital media experimentation, and hybrid forms of creative expression that combine AI-generated content with traditional artistic techniques.

6. **Benchmarking State-of-the-Art Models:** The dataset provides a benchmark for evaluating the performance of state-of-the-art AI models in real vs. fake image classification. It supports both supervised and unsupervised learning techniques, facilitating further development of more accurate and efficient models for detecting AI-generated content in real-world applications.

7. **Encouraging Innovation in Media Integrity Solutions:** The availability of this dataset encourages further research into the development of advanced detection mechanisms that go beyond traditional image recognition techniques. By providing both AI-generated and real-world images, it challenges researchers to develop innovative solutions that can detect subtle discrepancies in AI-generated images, ensuring greater media integrity across digital platforms.

Model SelectIon:

In our work, we conducted an extensive comparison of some deep learning models to IdentIfy AI-generated versus real art from a custom-created dataset.We started by utilizing strong pre-trained convolutional neural networks such as **ResNet50, VGG16, DenseNet121, EfficientNetB0, XceptIon, and InceptIonV3**. These models Are chosen due to their proven performance In Image classIficatIon and are adapted through transfer learnIng by removIng their top layers and addIng custom fully connected layers suitable for binary classifIcatIon. We also designed a **custom CNN model from scratch**, stacking convolutional blocks wIth batch normalizatIon, pooling, and dropout for regularIzatIon, wIth the aim of learning features specific to our dataset.

To train, we utilized data augmentation over Images to provide better dataset generalizability and diversity. Each model was both trained and validated based on a shared ImageDataGenerator pIpeline. Upon trainIng completion, we preserved all models and performed tests upon them against a hold-out, separate validation set with the measure of significant performance metrics such as accuracy, precisIon, recall, F1-score, and confusion matrices.

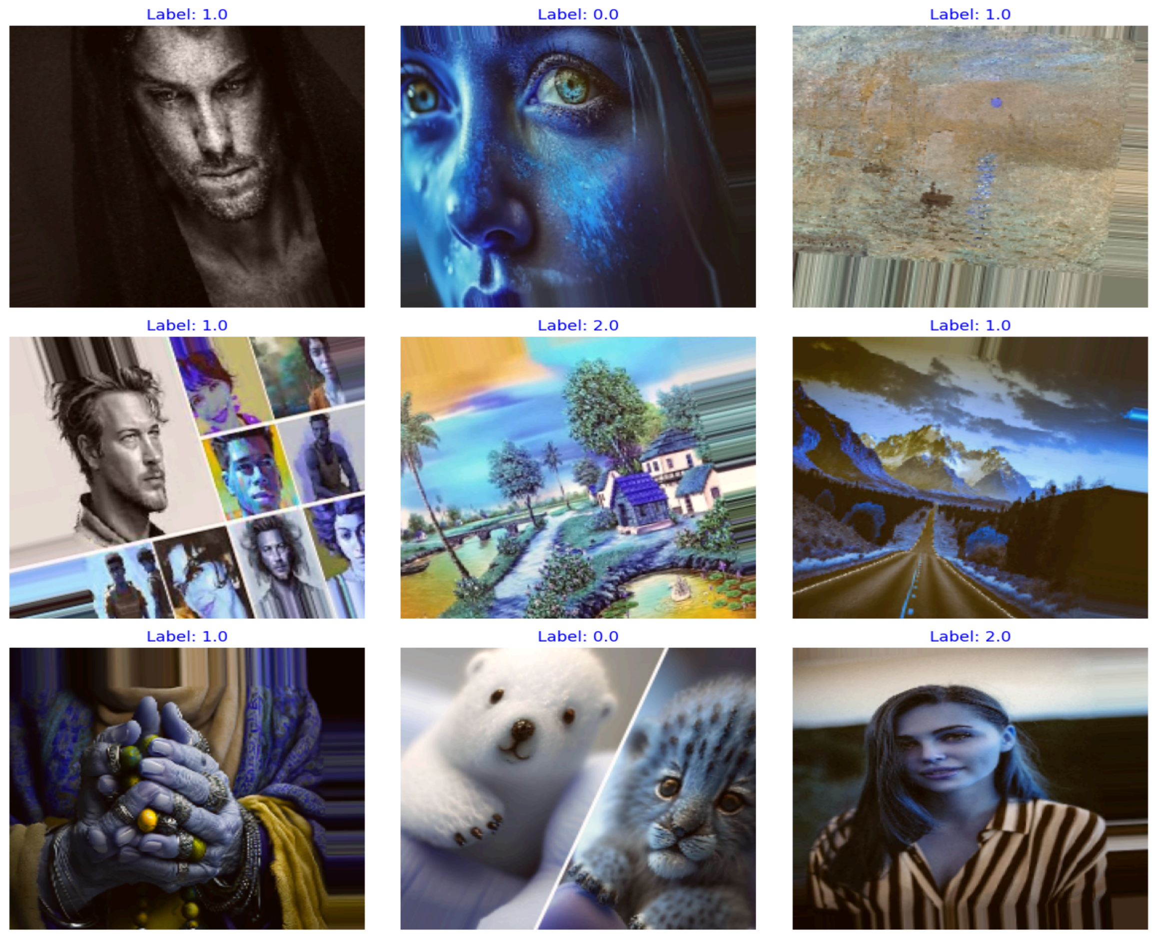
For comparison, we loaded the trained models and made predictions on the validation dataset. We used **classificatIon_report** and **confusion_matrix** from **ScIkit-learn to quantify performance**. The results are tabulated and vIsualized using confusion matrix heatmaps In order to get a better Insight Into where each model did It or not. ThIs approach provided a holistic assessment, allowing us to determine which model was best for the specific nature of the AI vs. real art classification task based on quantitative results and visual inspection.

Data Pre-processing:

For the **Custom CNN and ResNet50 models**, preprocessing ensures that the Input data Is transformed Into a suitable form that can enhance the learning ability of the models. The first preprocessing step Is to rescale the For the **Custom CNN and ResNet50 models,** preprocessing is essential to transform the input data into a suitable format that enhances the models' learning capabilities. The first step in preprocessing is to rescale the pixel values of the images, typically to a range of 0 to 1. This normalization improves the model's training process. For the **ResNet50 model,** this normalization is done using the **`preprocess_input`** function from TensorFlow, which is specifically designed for the ResNet architecture.

Data augmentation is another crucial preprocessing technique to enrich the training dataset, enabling the models to generalize better. By applying random transformations such as rotation, zoom, shifting, and flipping, the model learns to recognize features from various angles and sizes, which helps alleviate overfitting and enhances robustness.

Additionally, the dataset is split into a training set and a validation set, with **20%** allocated for validation. This allows the model to be tested on unseen data during training. The images are resized to a standard size, for example, 224x224 pixels, to ensure consistent input dimensionality, which is vital when using pre-trained models like ResNet50. This preprocessing pipeline guarantees that the models receive high-quality input data, enabling them to capture important features and improve classification accuracy.the pixel value of the Image, typically in a range of 0 to 1, to normalize the data and Improve the model training. For the **ResNet50** model, this Is achieved through the **preprocess_Input** function of TensorFlow, which Is specifically designed for the ResNet architecture. Data augmentation Is another vital preprocessing technique employed to Improve the richness of the training data set to allow the models to generalize better. Through the application of random transformations such as rotation, zoom, shift, and flipping, the model learns to recognize features at various angles and sizes, alleviating overfitting and enhancing robustness.
Additionally, the data Is split Into a training set and a validation set with a valIdatIon split of **20%** so that the model can be tested on unseen data while training. The Images are resized to a standard size, for example, 224x224 pIxels, to ensure Input dimensionality consistency, which Is essential when using pre-trained models like ResNet50. This preprocessing process ensures that the models receive high-quality Input data so they can capture good features and Improve classIfIcatIon accuracy.

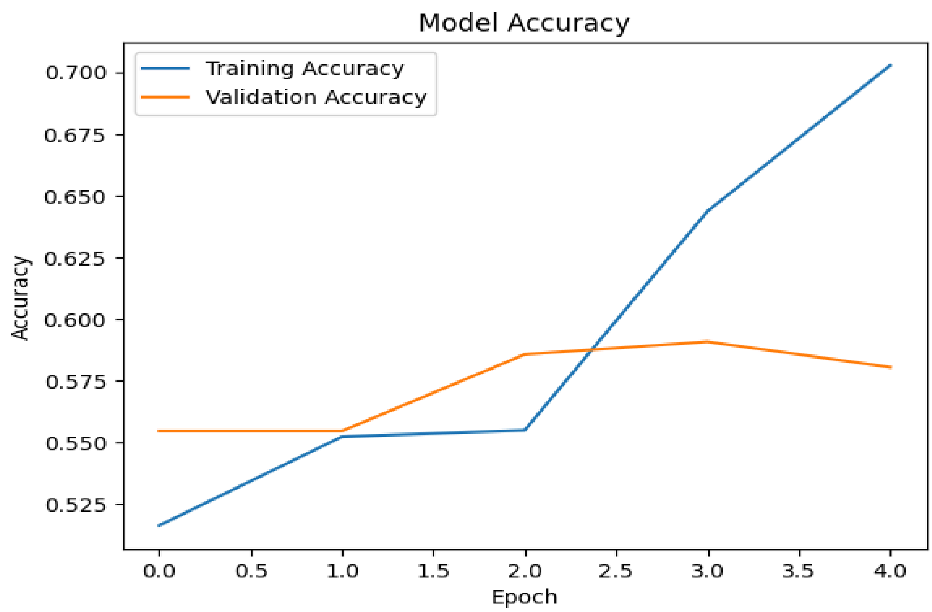


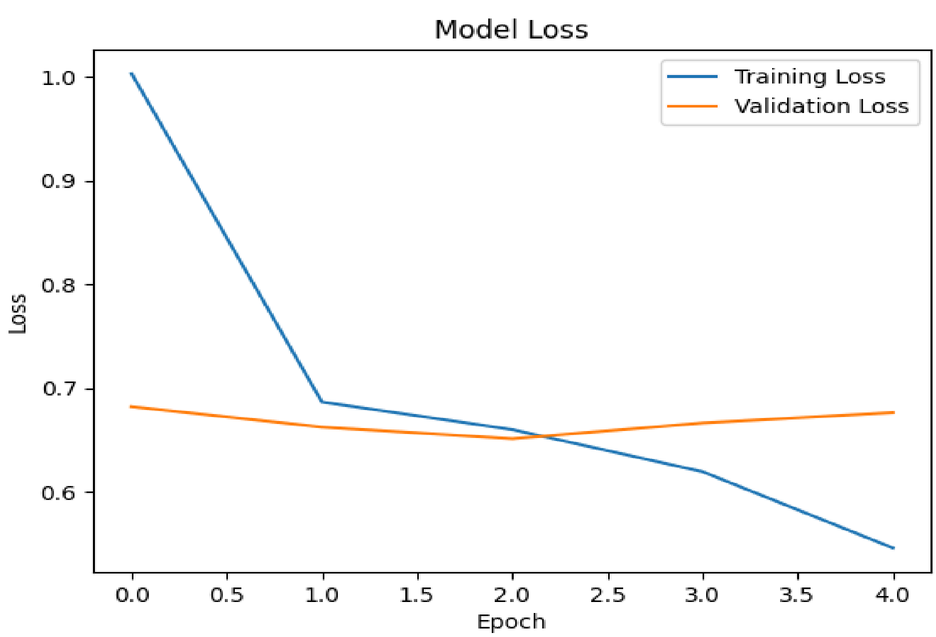
Train Test Splitting:

In this study, the data set consisting of AI-generated art and real artwork Images was split Into a training set and a validation set using the **`ImageDataGenerator`** class In TensorFlow. Specifically, an **80-20** splIt was performed by reserving **80%** of the data for training and the remaining **20%** for valIdatIon. This was achieved using the **`valIdatIon_splIt`** argument and the `subset` parameter of the **`flow_from_dIrectory()`** method. The training subset was used to acquire the visual features of each class to train the model, and the validation subset was used to check how the model works on unseen data during training In an attempt to prevent overfitting.

TraIn-test splitting Is highly crucial In machine learning, especially for Image class tasks. It. Ensures that the model Is not memorizing the training data but learns to generalize. Features and Is able to work on new, unseen Instances. It Is Possible to monitor In real time the accuracy and loss of the model using a valIdatIon set for training and apply early stopping or checkpointing policies based on valIdatIon performance. This approach enhances the reliability of the model and Its evaluation measures to better reflect Its performance In the real world, rather than Its success on the training data. This train-validation approach Is especially critical when there are Imbalanced or small training sets because It prevents the model from taking over the estimation of its generalizatIon abIlIty.

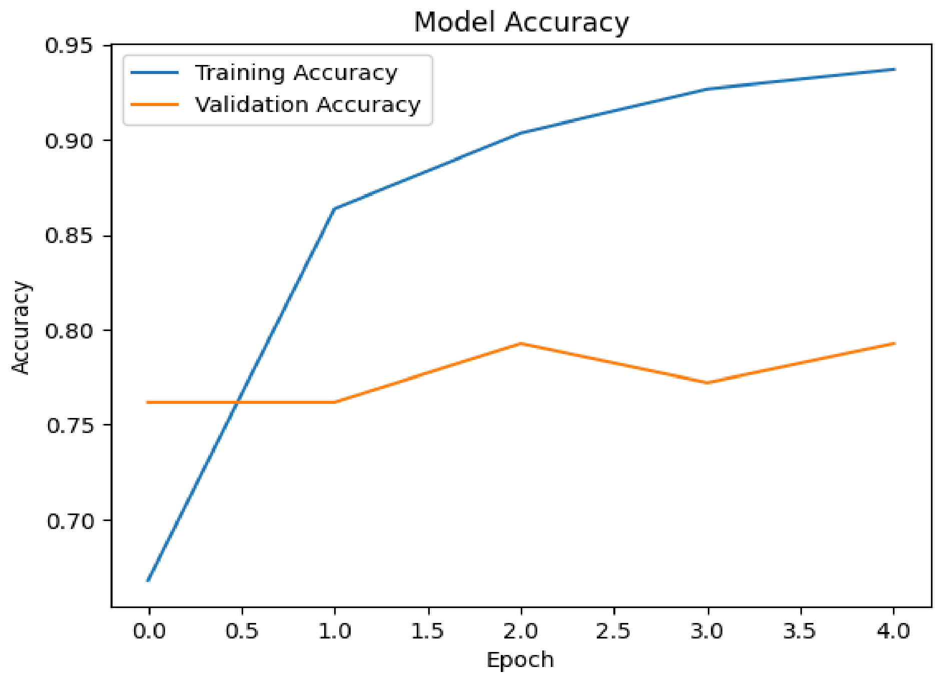
VGG16 Reports:

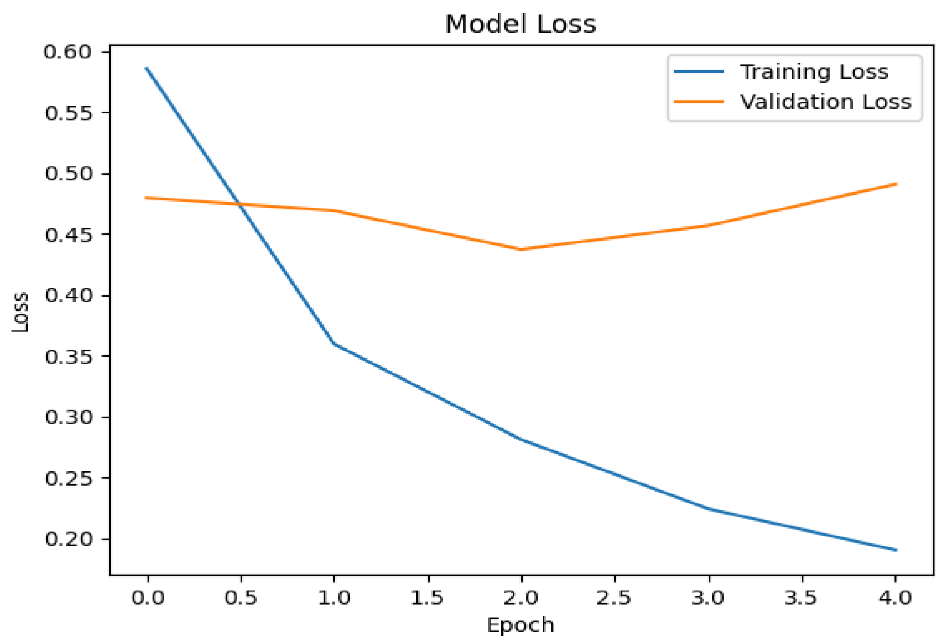




Validation Accuracy: 58.03%

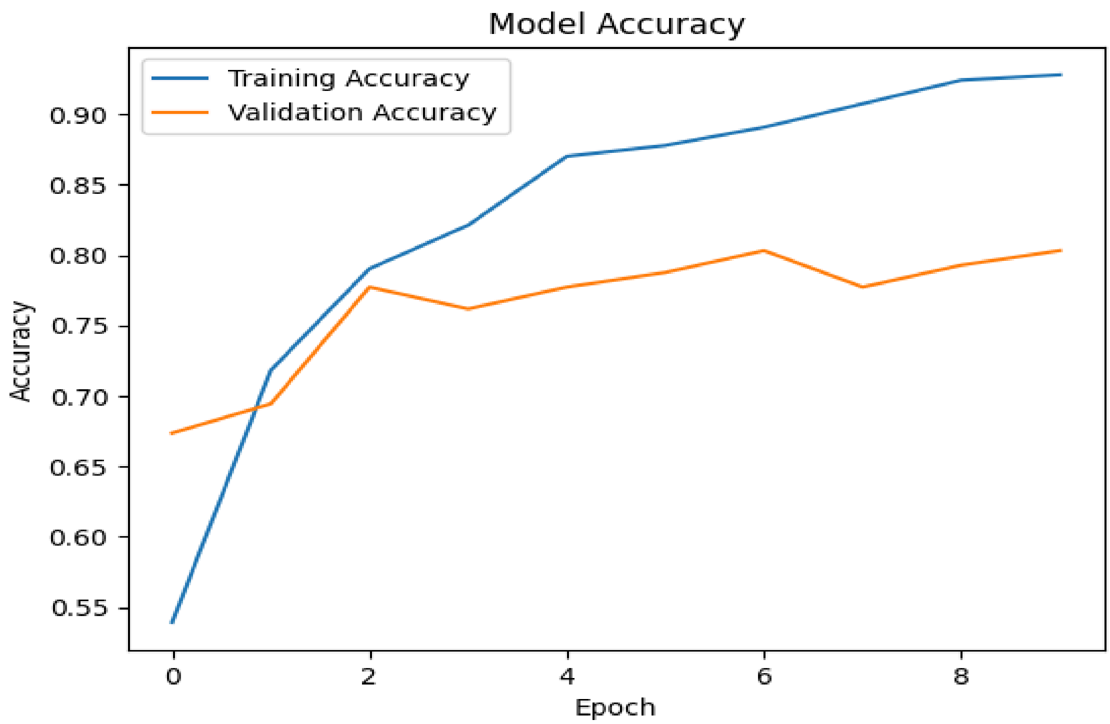
EfficientNetB0 Reports:

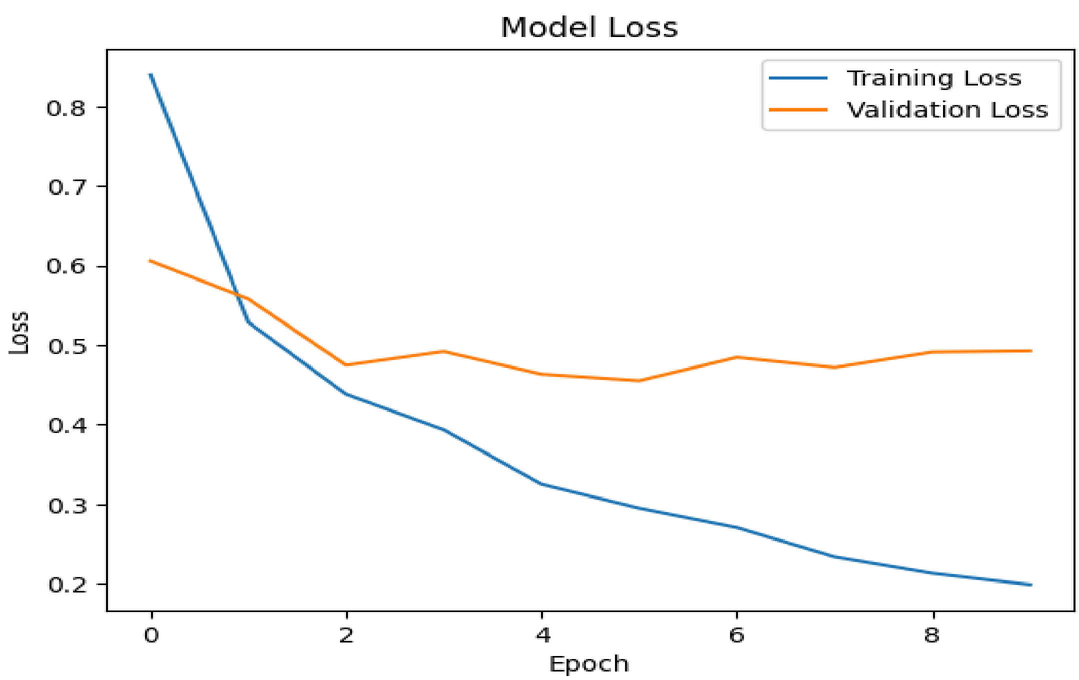




Validation Accuracy: 79.27%

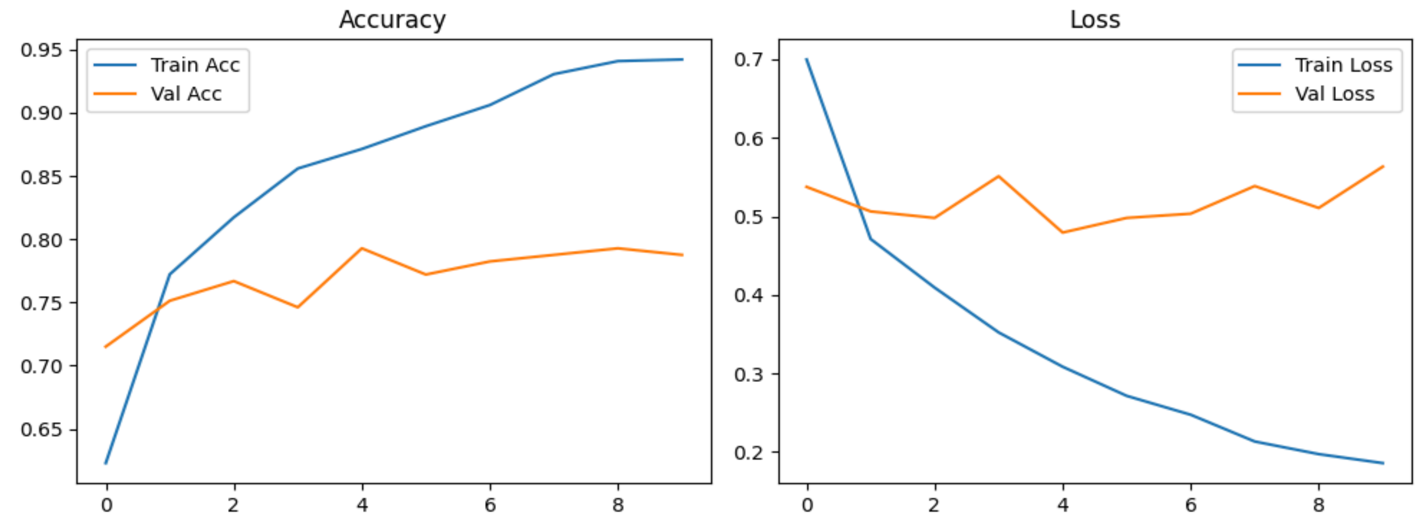
Inception V3 Reports:





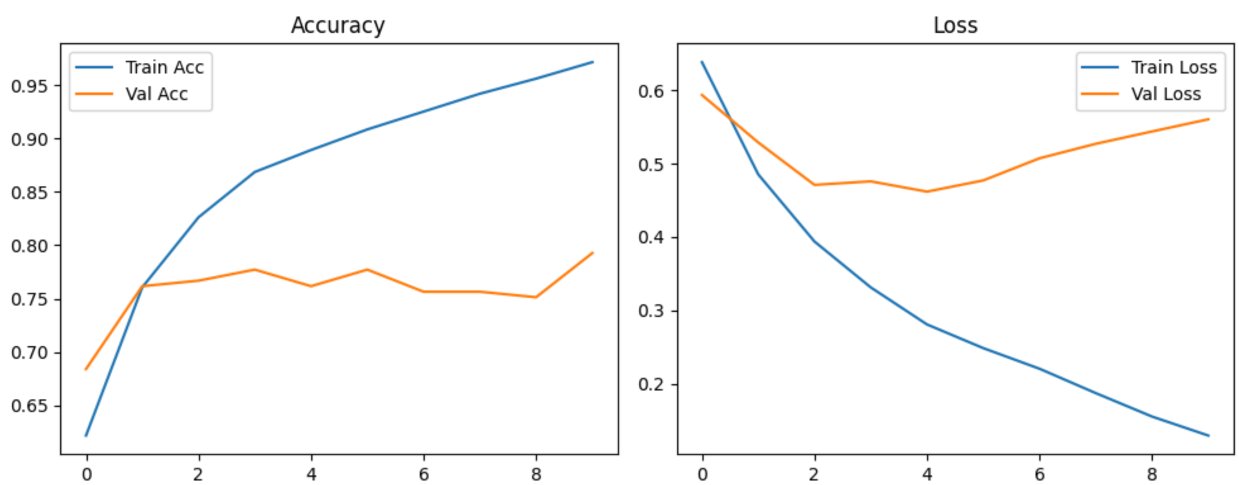
Validation Accuracy: 80.31%

DenseNet121 Reports:



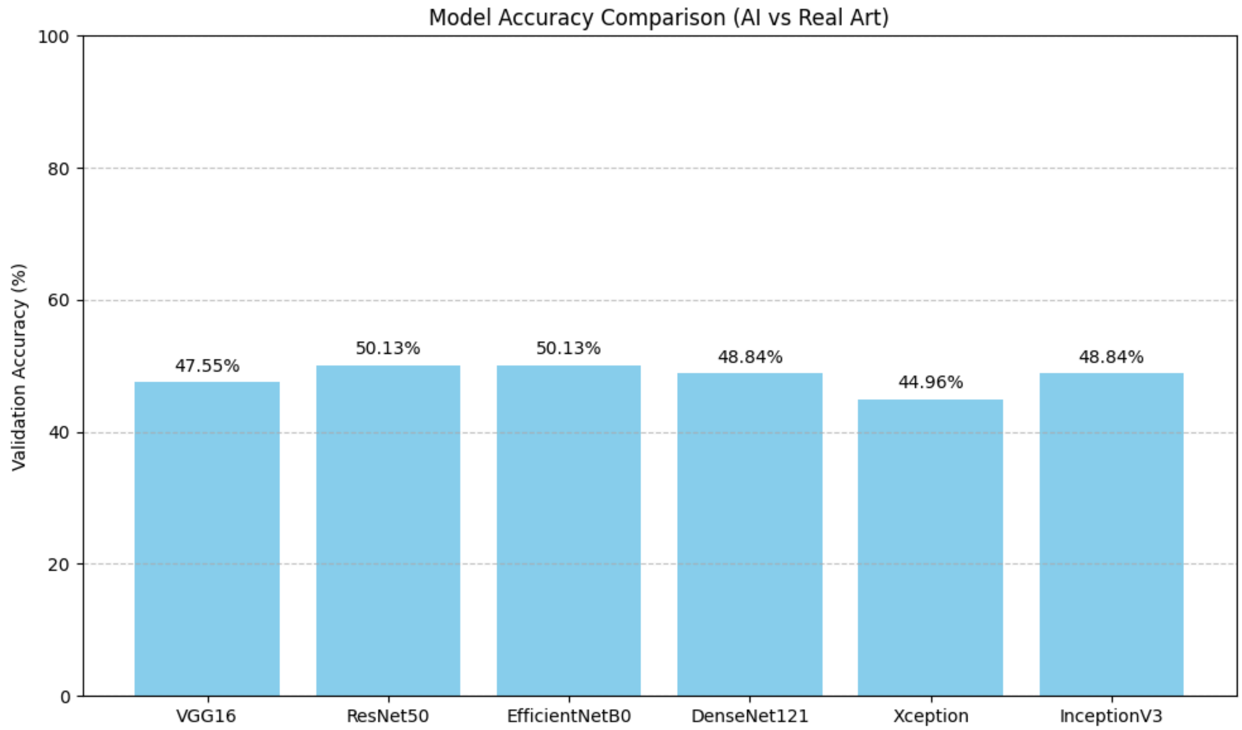
Validation Accuracy: 78.76%

Xception Reports:



Validation Accuracy: 79.27%

All Model Comparison Report:

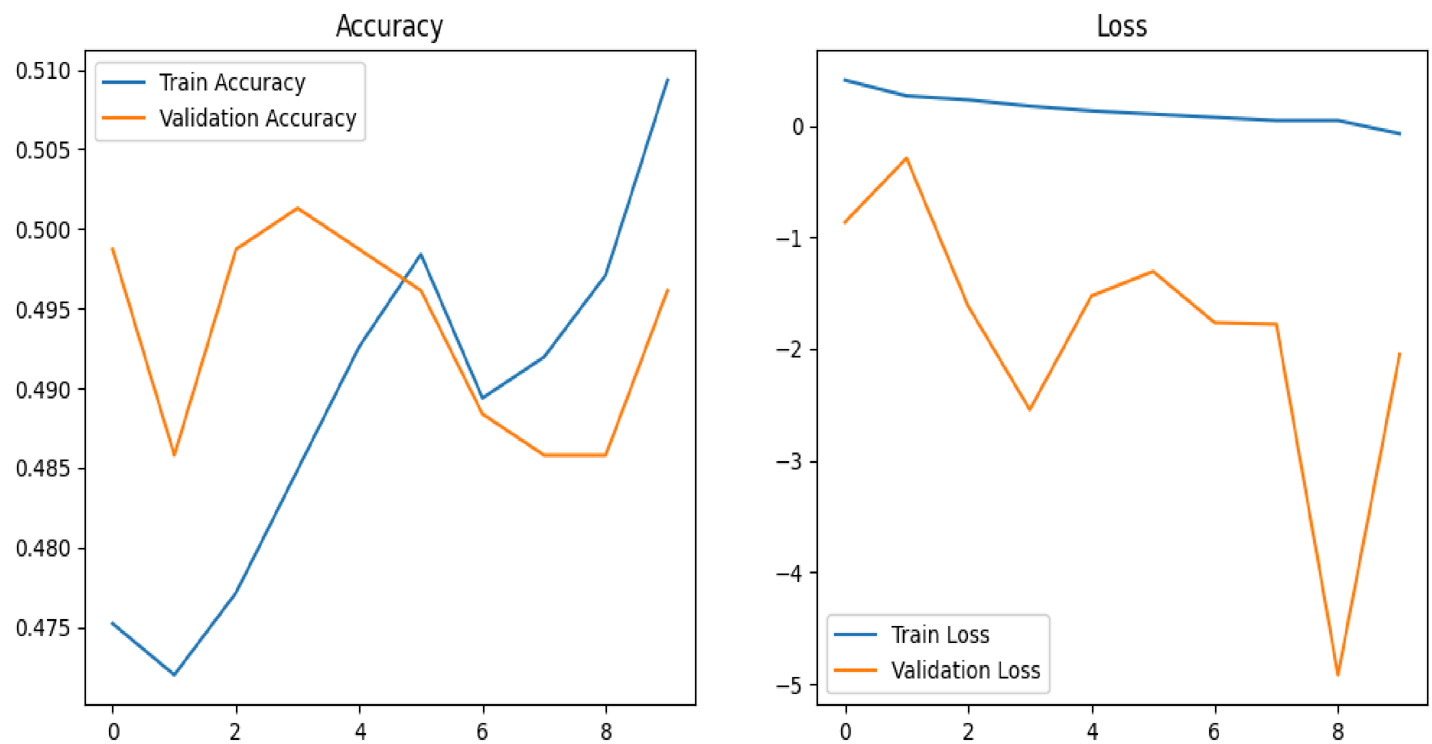


Best Model: ResNet50 with Accuracy: 50.13%

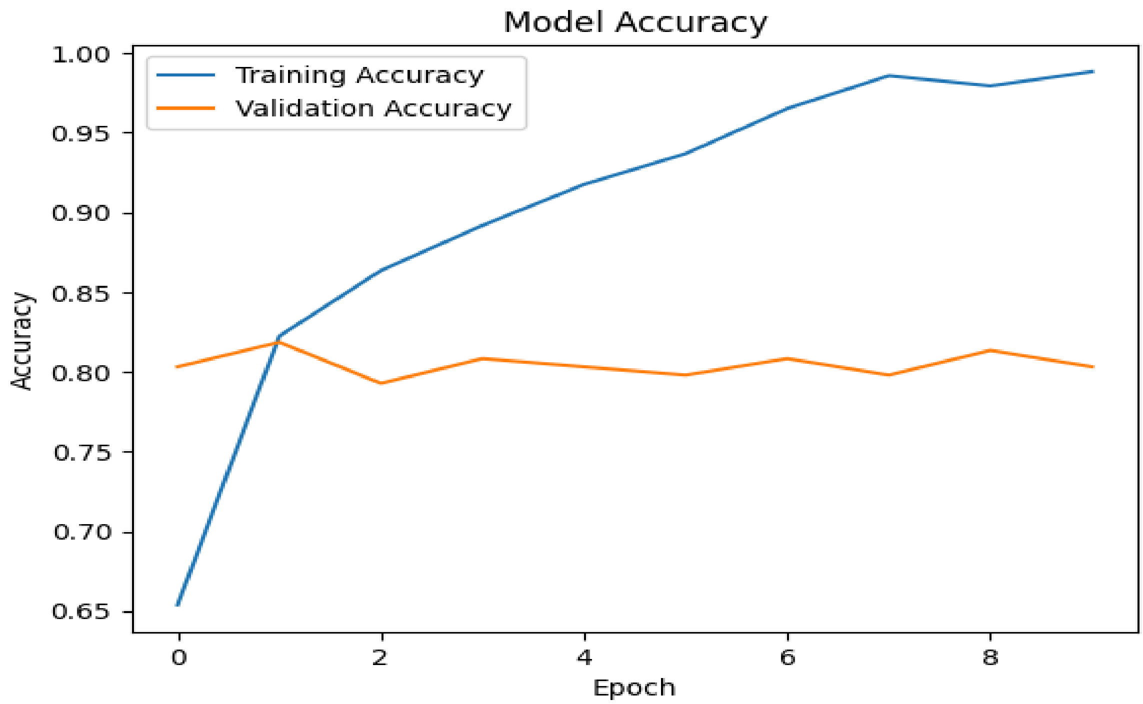
Results And Discussion:

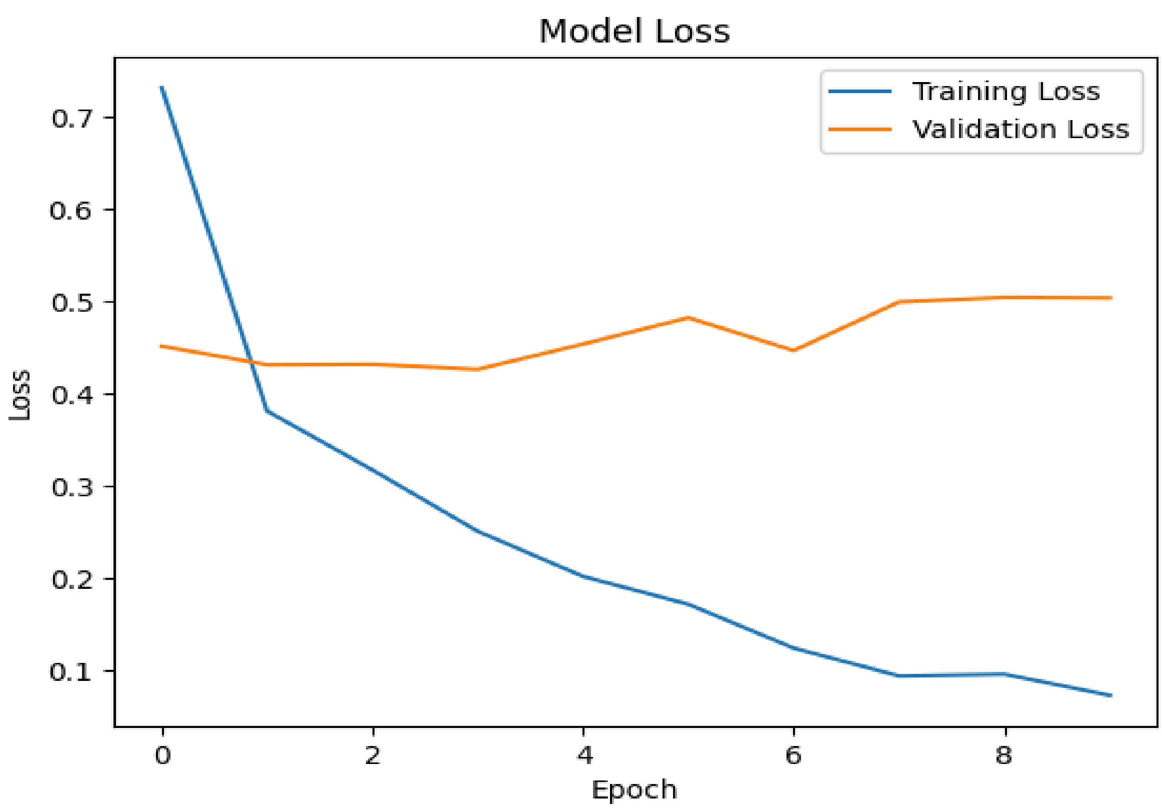
The **comparison of the ResNet50 model and Custom CNN model** brings to light several sIgnIfIcant poInts. The **ResNet50** model performs better than the **Custom CNN** with **57%** accuracy against the Custom CNN's **46%**. This difference illustrates the advantage of using a pre-trained model such as **ResNet50**, which benefits from learned features from **ImageNet** and generalIzes more effectively to new datasets. However, both models suffer from class Imbalance because they are more accurate for predicting the majority class (RealArt) compared to the minority class (AIArt). The Custom CNN model Is extremely accurate for **AIArt (0.67)** but Is very low on recall **(0.04)**, and the ResNet50 model Is very high In recall for **AIArt (0.97)** but Is low In precision for **RealArt (0.07).** Both models are based on the majority class, and the reason Is most likely an unbalanced dataset. The Increased accuracy and higher recall of ResNet50 for AIArt suggest that It Is superior In feature extraction and generalIzatIon than Custom CNN. Neither of the models, however, performs Ill, suggesting that further enhancements In hyperparameter tuning, data augmentation, and class balancing are required. Generally, although **ResNet50** yields superior results, the two models must be optimized to cope with class Imbalance and Increase performance.

This Image represents the reports of the custom CNN model that we created:

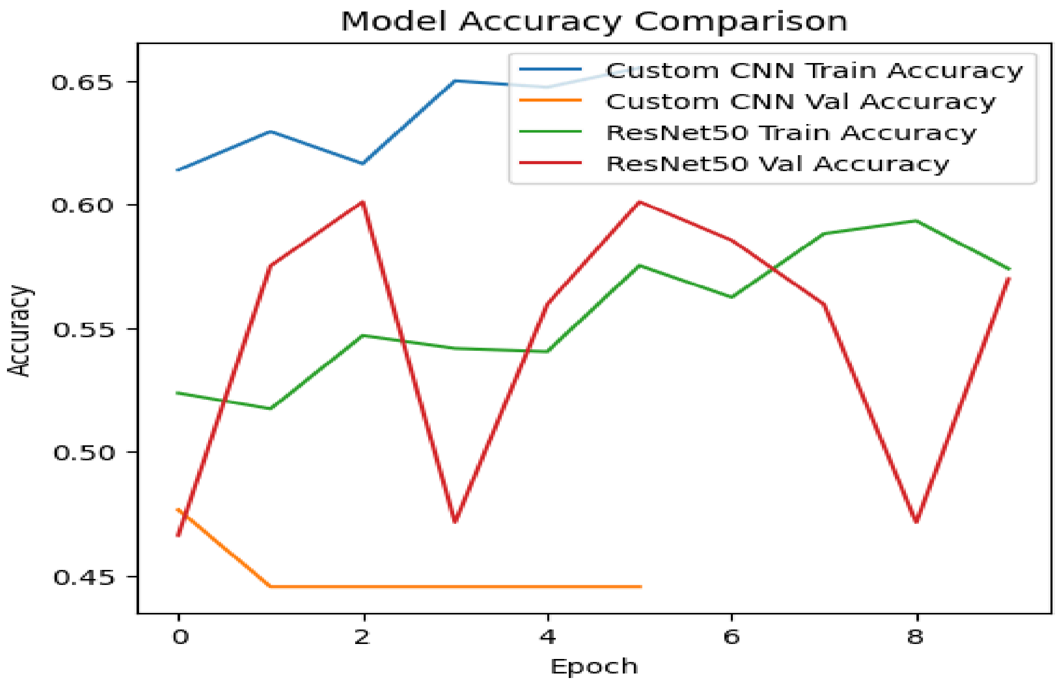


This Image represents the ResNet50 reports:



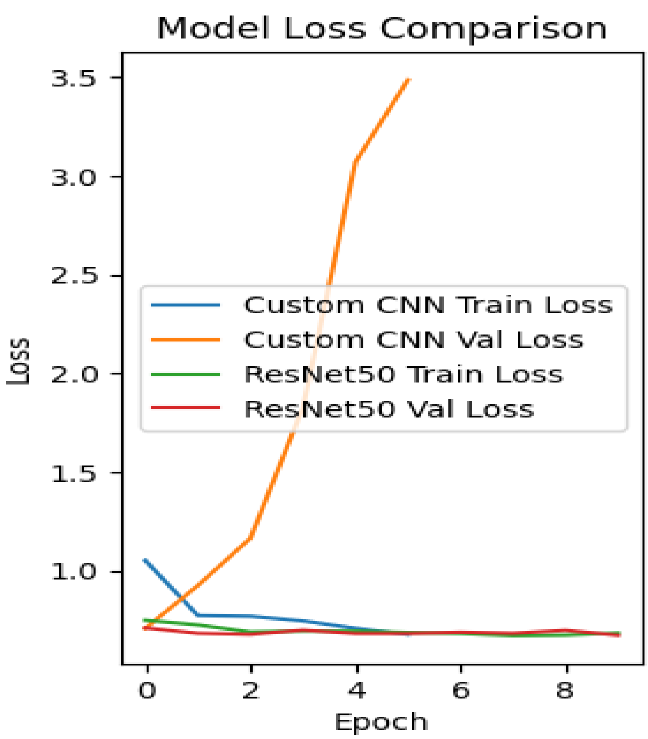


Plot Accuracy Comparison:



ResNet50 ValIdatIon Accuracy: 59.07% Custom CNN ValIdatIon Accuracy: 43.01%

Model Loss Comparison:



Future Work:

For future studies, different techniques can be used to improve the performance of both the ResNet50 and Custom CNN models. Handling class Imbalance Is one area of Improvement because both models perform better when dealing with the majority class (RealArt). Techniques like oversampling the mInorIty class (AIArt), undersampling the majority class, or using class LIghts In the loss function would be capable of reducing this and enhancing fairness and performance on both classes. Additionally, expanding the training dataset with more advanced data augmentation techniques like random cropping, brightness/contrast adjustment, or mIxup could enhance the models' general information further. Hyperparameter tuning also needs to be enhanced; experimenting with optImal learnIng rates, batch sizes, and layer numbers using grid search or random search could yield Improved results. Though ResNet50 has achieved satisfactory performance due to pre-traInIng, freezIng some of the layers and fIne-tunIng them on the destination dataset can give even better accuracy and recall, particularly for the minority class. Experimenting with more complex model structures, such as EffIcIentNet, VGG16, or Transformer models, might yield new findings and further gains. Further, experimentation to merge both models Into an ensemble strategy such as stacking could leverage theIr strengths and further Improve overall performance. Lastly, employing metrics beyond accuracy, such as F1-score, PrecIsIon-Recall AUC, and ROC AUC, would give a clearer picture of the performance of the models, especially on unbalanced datasets. By paying attention to these features, subsequent studies can lead to more balanced, precise, and stable models that are both good for the majority and minority classes, as well as exploring other data preprocessing and model topics to reach even higher class performance.

References:

Goodfellow, I. et al. "Generative Adversarial Nets," NIPS, 2014.

Kingma, D. P., & Welling, M. "Auto-Encoding Variational Bayes," ICLR, 2014.

Rombach, R. et al. "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR, 2022.

Ramesh, A. et al. "Zero-Shot Text-to-Image Generation," OpenAI, 2021.

Radford, A. et al. "Learning Transferable Visual Models From Natural Language Supervision," ICML, 2021.

Rombach, R. et al. "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR, 2022.

Matern, F. et al. "DeepFake Detection: A Survey," 2020.

Zhou, X. et al. "Fake Image Detection Using CNNs," 2021.

Shao, X. et al. "Detection of AI-Generated Images: A Comprehensive Review," 2022.

Dolhansky, B. et al. "The DeepFake Detection Challenge Dataset," CVPR, 2020.

Rossler, A. et al. "FaceForensics++: Learning to Detect Manipulated Facial Images," CVPR, 2020.

Li, X. et al. "Fake Image Detection Dataset," 2023.

Karras, T. et al. "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR, 2019.

Brock, A. et al. "BigGAN: Large Scale GANs for Image Synthesis," ICLR, 2019.

Razavi, A. et al. "VQ-VAE-2: Generating Diverse High-Fidelity Images with VQ-VAE-2," NeurIPS, 2019.

Liu, Y. et al. "Ensemble Methods for Fake Image Detection," 2021.

Pan, S. et al. "Transfer Learning for Fake Image Detection," 2022.

Made By:

Tanvir Manhmud Ove,
Department Of Software Engineering,
Metropolitan University.