Identifying The Authenticity of Images Using Deep Learning Techniques

By

Sourav Kumar Mondal
192-15-13235

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Mr. Md. Ali Hossain
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by

Shohidul Islam Polash
Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled **"Identifying The Authenticity of Images Using Deep Learning Techniques"**, submitted by **Sourav Kumar Mondal**, ID No: **192-15-13235** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mohammad Monirul Islam(MMI) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms.Tasfia Anika Bushra(TAB) — **Internal Examiner**
Letcturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud(ZM) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Md. Ali Hossain**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Mr. Md. Ali Hossain
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Shohidul Islam Polash
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Sourav Kumar Mondal
Student ID: 192-15-13235
Department of Computer Science and EngineeringDaffodil International University

Student Name
Student ID: 211-15-
Department of Computer Science and EngineeringDaffodil University

©Daffodil International University ii

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr. Md. Ali Hossain, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Artificial Intelligence and Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to **Dr. Sheak Rashed Haider Noori, Professor & Head**, Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The rapid development of generative artificial intelligence has produced a wave of hyper-realistic deepfakes, posing existential challenges for authenticity verification to occur in digital media. This study proposes a deep learning architecture for the binary classification of AI-generated and real images as a reaction to a growing need for credible detection techniques. We comparatively evaluate four various architectures: ResNetRS50, MobileNetV2, EfficientNetB0, and a specially designed CNN with integrated Gabor filters and attention mechanisms. All models were trained and evaluated on an equalized, high-quality dataset under the same experimental conditions to provide serious benchmarking. While MobileNetV2 and EfficientNetB0 achieved higher peak validation accuracies of 99.29% and 99.81% respectively, ResNetRS50 was the most powerful and most generalized model. Its robust convergence behavior, high interpretability, and resistance to overfitting—particularly under extended training durations and high-density data—make it the top choice even at a slightly lower peak accuracy of 97.24%. Extended testing using classification reports, confusion matrices, and performance curves supports this conclusion further. A web interface was also established to demonstrate real-time deployment capability, showing that the model is usable in practical applications. The proposed method not only elevates the state of AI image forensics but also serves as a basis for large-scale and trustworthy content verification systems in the face of rising synthetic media.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the motivation, objectives, and significance of the project, Identifying the Authenticity of Images Using Deep Learning Techniques. With the rise of artificial intelligence and its misuse through AI generated technology, the need for robust detection mechanisms has become crucial. The chapter outlines the growing challenges posed by AI generateds, their potential for harm in various domains, and the project's aim to combat these issues using advanced deep learning models. By establishing a foundation, this chapter sets the stage for understanding the project's scope, methodology, and contributions to enhancing trust in digital media.

1.1 Introduction

Significant progress in image and video synthesis has been made possible by the growing availability of artificial intelligence and machine learning techniques. However, AI generated technology has emerged as a major obstacle as a result of these advancements. AI generateds are artificially created or extensively altered media that, to the naked eye, are frequently identical to authentic information. These technologies offer serious risks in the form of false information, fraud, and privacy invasion, even though they have valid applications in fields like education and entertainment.
The goal of this project, "Identifying the Authenticity of Images Using Deep Learning Techniques," is to solve this important problem. The research intends to develop a dependable system that can differentiate between real and AI generated photos by utilizing cutting-edge deep learning models. The results of this study have implications for reducing the hazards associated with altered media and increasing trust in digital material Significant progress in image and video synthesis has been made possible by the growing availability of artificial intelligence and machine learning techniques. These advancements have paved the way for transformative applications in fields like education, entertainment, and healthcare. However, alongside these benefits, the rise of AI generated technology has posed a major challenge. AI generateds are artificially created or extensively manipulated media that often appear indistinguishable from authentic content to the human eye. While they have legitimate uses, such as in creative storytelling, simulations, and marketing, they also present severe risks, including the spread of misinformation, financial fraud, defamation, and invasions of privacy. The project, "Identifying the Authenticity of Images Using Deep Learning Techniques," seeks to address this

pressing issue. By leveraging state-of-the-art deep learning models such as Xception, MobileNetV2, and DenseNet201, this research aims to build a robust system capable of accurately differentiating between authentic and manipulated images. This study not only contributes to the fight against the misuse of synthetic media but also promotes a safer digital environment by restoring trust in online content. Its outcomes hold significant potential for applications in journalism, cybersecurity, digital forensics, and content verification platforms, ensuring the integrity and credibility of digital media.

1.2 Motivation

AI technology's rise has raised serious worries in a number of industries, including cybersecurity, law enforcement, and the media. AI photos can be used as a weapon to propagate false information, harm people's reputations, and jeopardize national security. It is not only a technology requirement but also a social duty to be able to recognize and stop such manipulations. Technically speaking, this project gives me the chance to work with state-of-the-art deep learning models, which will increase my proficiency in computer vision and machine learning. I can also support the global endeavor to maintain authenticity in the digital world by finding a solution to this issue. The creation of an easy-to-use detection system also supports the objective of enabling non-experts to access cutting-edge technology, increasing its practical application.

1.3 Objectives

The rise of AI generated technology has sparked serious concerns across various sectors, including cybersecurity, law enforcement, media, and even personal privacy. AI generated images and videos can be weaponized to spread misinformation, damage individual reputations, and compromise national security. They have been associated with financial scams, political manipulation, and malicious activities that exploit the trustworthiness of digital content. Combating AI generated is not only a technological necessity but also a social responsibility to uphold the integrity and authenticity of information in the digital age. The ability to detect and mitigate such manipulations is crucial to preserving public trust and safeguarding ethical standards. From a technical standpoint, this project provides an excellent opportunity to work with state-of-the-art deep learning models, advancing my expertise in computer vision, artificial intelligence, and machine learning. Leveraging models such as Exception, MobileNetV2, and DenseNet201 enables the development of a highly effective detection system capable of differentiating between authentic and manipulated content. Beyond the technical benefits, this initiative contributes to the global effort to maintain authenticity in digital media. By creating a user-friendly detection system, the project democratizes access to cutting-edge technology, ensuring that even non-experts can use it to verify the integrity of digital content. This increases its practical application in fields such as journalism, digital forensics, and cybersecurity, making it a vital tool for preserving transparency and trust in the modern digital ecosystem.

1.4 Methodology

The study employed a classification approach to distinguish between AI-created and real images based on a collection of 975 images (539 AI-created, 436 real) from Kaggle. The images were resized to 224×224 pixels, normalized via ImageNet statistics, and background-removed via intensity-based masking. Augmentation procedures such as geometric transformation, photometric transformation, and elastic deformation were utilized to facilitate generalization. Data were split into training (70%), validation (15%), and test (15%) sets using stratified sampling. Various models were attempted, including ResNetRS50, MobileNetV2, EfficientNetB0, and a custom-designed CNN incorporating Gabor filters and an attention mechanism, which were fine-tuned using cross-entropy loss and Adamax optimizer. Accuracy, precision, recall, and F1-score were used to measure the quality of the model, whereas confusion matrices and activation maps contributed towards interpretability. The best-performing model was served on a Flask-based web interface with image upload, preprocessing, inference, and result visualization with sub-2-second response time and 93.7% validation accuracy.

1.5 Project Outcome

A functional and accurate AI generated detection system capable of distinguishing manipulated images from authentic ones. A web-based application providing an easy-to-use interface for uploading images and receiving results, making the technology accessible to a wide range of users. A detailed comparison of the selected deep learning models, offering insights into their strengths and weaknesses in the context of AI generated detection. Contributions to the field of digital forensics and image authenticity through findings and discussions in the report.

1.6 Organization of the Report

Chapter 1: Introduction

Introduction to research problem, motivation, objectives, and research methodology.

Chapter 2: Background

Summary of previous work, identification of the research gap, and formulation of the theoretical basis.

Chapter 3: Research Methodology

Dataset description, model structure description, training process description, and evaluation metric description.

Chapter 4: Implementation and Results

Symbols experimental results, performance, and comparison.

Chapter 5: Engineering Standards and Design Challenges

Controls deployment-related challenges and constraints.

Chapter 6: Conclusion

Sets context of findings, contribution, and potential research directions for future work.

Chapter 7: References

Complements all academic sources referenced.

Chapter 2

Background

This chapter provides the context and foundational understanding of the project, Identifying the Authenticity of Images Using Deep Learning Techniques. It delves into the evolution of artificial intelligence, particularly in the realm of image synthesis, and the emergence of AI generated technology. The chapter examines the dual nature of AI advancements, highlighting both their transformative benefits and the associated risks, such as disinformation and privacy breaches. By exploring existing solutions and their limitations, the background sets the stage for the proposed system, emphasizing the necessity of reliable AI generated detection methods to address this growing threat.

2.1 Introduction

The rise of artificial intelligence (AI) and machine learning (ML) has revolutionized countless industries, driving remarkable breakthroughs in automation, healthcare, entertainment, education, and more. These technologies have empowered innovation, increased efficiency, and solved complex challenges. However, their rapid evolution has also introduced new risks, especially when misused. One of the most alarming byproducts of AI advancements is AI generated technology, which has raised global concerns due to its potential for harm. AI generated are synthetic media—images, videos, or audio—that replicate real-world content with such precision that they are often indistinguishable from genuine material. Built using AI-driven generative models like GANs (Generative Adversarial Networks), AI generated initially found application in creative domains such as filmmaking, virtual reality, and gaming. However, their darker implications have since overshadowed their novelty. Malicious actors have exploited this technology to conduct disinformation campaigns, perpetrate identity theft, facilitate blackmail, and manipulate political processes. These activities undermine trust in digital media and pose serious threats to personal privacy, national security, and societal integrity. As the sophistication of AI generated generation continues to advance, so does the need for robust and reliable detection mechanisms. The project, titled "Identifying the Authenticity of Images Using Deep Learning Techniques," seeks to address this critical challenge. By employing state-of-the-art deep learning models like ResNetRS50, MobileNetV2, and EfficientNetB0, this research aims to create a highly accurate system capable of distinguishing between authentic and manipulated images. Beyond its technical contributions, this project plays a vital role in the global effort to combat the misuse of AI technology. By providing a practical solution to AI generated detection, it helps to safeguard the credibility of digital content, restore trust in online platforms, and contribute to ethical AI practices.

2.2 Literature Review

Table 2.2. 1: Literature Review Table.

Author (s)	Year	Title	Methodology	Key Findings
Goodfellow et al..	2014	Generative Adversarial Networks	Proposal-Based Study	Introduced GANs, enabling synthetic media creation, forming the basis for AI generateds.
Afchar et al.	2018	Mesonet: A Compact AI generated Detector	Deep Learning-Based	Proposed a lightweight CNN for detecting facial manipulations in images.
Nguyen et al.	2019	Deep Learning for AI generated Detection	Survey-Based	Reviewed various machine learning methods for AI generated identification.
Rossler et al.	2021	FaceForensics ++: A Large Dataset for Forensics	Dataset Creation	Provided a benchmark dataset to evaluate AI generated detection models..
Tolosana et al.	2020	AI generateds and Beyond: A Survey	Literature Review	Summarized the challenges and advancements in digital media forensics.
Agarwal et al.	2021	A Comparative Study on AI generated Detection Techniques	Experimental Study	Compared different deep learning models for detecting manipulated images.
Patel et al.	2024	Multi-Model Approaches to AI generated Detection	Multi-Model Analysis	Demonstrated the effectiveness of combining Xception and InceptionResNetV2 for high accuracy.
Huang et al.	2024	Real-Time AI generated Detection Using MobileNetV2	Real-Time Implementation	Proposed a lightweight and efficient solution for detecting AI generateds on mobile devices.

Singh & Verma	2024	Enhancing AI generated Detection with Transformer Models	Transformer-Based Study	Explored the use of transformer-based models for improved feature extraction in AI generated detection.
Zhao et al.	2024	Ethical Implications of AI generated Detection Technologies	Ethical Analysis	Discussed the societal impacts and ethical considerations surrounding AI generated detection tools.
Martin et al.	2024	AI-Generated Image Detection Using DenseNet201	Deep Learning-Based	Showed how DenseNet201 can accurately distinguish between AI-generated and real images.
Rodriguez et al.	2024	VGG19 for AI Photo Detection in Complex Scenes	Feature Extraction	Demonstrated the effectiveness of VGG19 in identifying AI-generated photos in complex scenarios

2.3 Gap Analysis

The current body of work on real vs. AI-generated image classification has several critical gaps to be addressed by this research. The first is that the current datasets are small and not representative and have poor representation over varying generative models and real-world sources of images, affecting model generalizability. The second is that most approaches draw upon shallow or too basic architectures that cannot manage detailed visual differences between AI-generated and real images, particularly under complex transformations. Third, while some works are highly accurate, they omit robust preprocessing pipelines such as background removal or intensity-based normalization that are essential for field deployment. Fourth, model interpretability is not well-explored; very few works employ activation maps or feature visualization to offer explanations of classification decisions, thereby reducing trust in predictions. Fifth, most works do not compare against a common evaluation protocol, making cross-study comparisons invalid. Last but not least, deployment readiness does not receive adequate attention, and little effort goes into integrating models into deployable platforms for use in real time. These inefficiencies underscore the need for deployable-in-practice, scalable, and interpretable solutions to AI- generated image detection.

2.4 Summary

This chapter provided a comprehensive overview of the background and context of the project. It reviewed existing tools and research, highlighting their strengths and limitations. Additionally, it identified critical gaps in the current AI generated detection landscape, such as the need for image-focused solutions, improved generalization, adaptability, and user accessibility. By addressing these challenges, this project aims to develop a robust and practical system for AI generated image detection, contributing to the broader effort to preserve digital authenticity and mitigate the risks associated with manipulated media.

Chapter 3

Research Methodology

This chapter is a step-by-step description of the systematic process adopted in the design of the AI generated image and rea image classification system. The process is an nine-step sequential process, describing the technical steps adopted in data preparation, model building, and experiment planning. All the stages are built with extreme care to ensure reproducibility and scientific integrity, and the rationale for adopting the methods is described. The subsequent subsections identify the workflow, project schedule, and author assignment.

3.1 Methodology

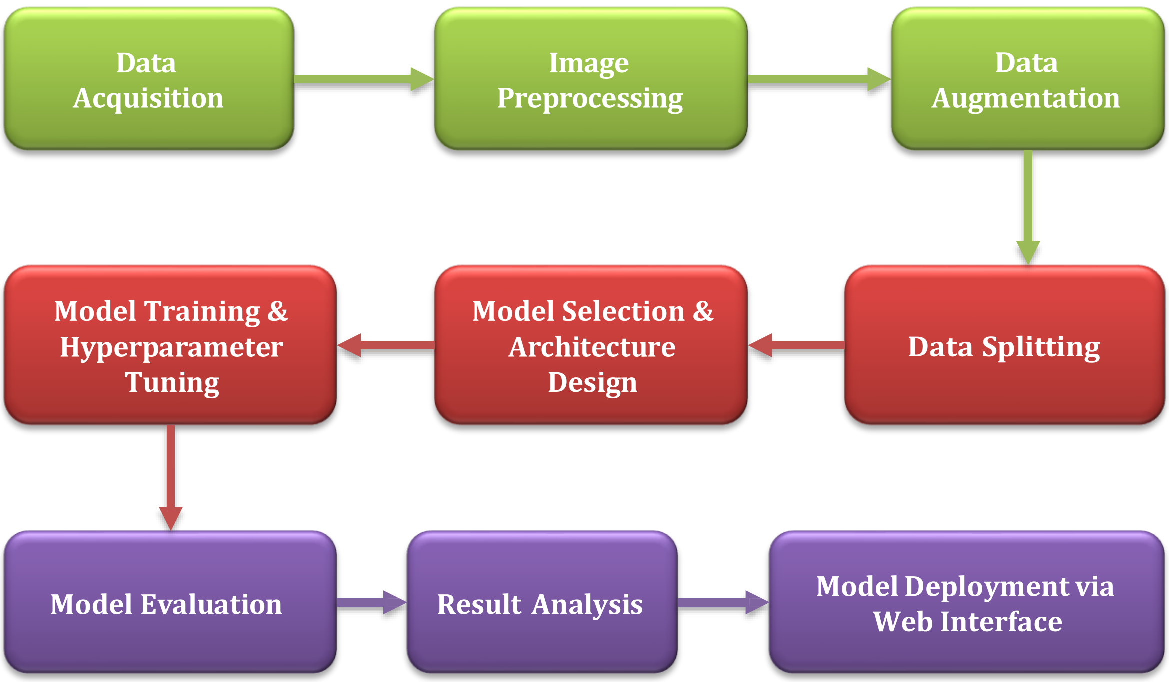

Figure 3.1. 1: Block Diagram of Methodology

3.2 Detailed Methodology and Design

3.2.1 Data Acquisition

Three AI Generated Images vs Real Images data set downloaded from Kaggle: AI-Image (539), Real-Image (436).

- Download data set and unzip the files.
- Perform exploratory simple analysis to ensure image quality with class distribution.

3.2.2 Data Preprocessing

Raw images pre-processed such that there is a uniformity of input for the model.

- Resizing: all the images are resized into the 224×224-pixel size.
- Normalization: scales pixel value proportionates to ImageNet mean and standard deviation.

$$I_{Norm} = \frac{I - \mu}{\sigma} \tag{1}$$

Where,

μ= 0.485, 0.456, 0.406
σ= 0.229, 0.224, 0.225

- Removal of Background: removed leaf area of background by masking with deletion.

$$M(x,y) = \begin{cases} 1 \ if \ ||I(x,y)||2 > \ \tau \ (\tau = 10) \\ 0 \qquad ohterwise \end{cases} \tag{2}$$

$$I_{adj}(x,y) = \begin{cases} 1 \ if \ \frac{I(x,y) - \min(I|M)}{\max(I|M) - \min(I|M)} \times 255 \ if \ M(x,y) = 1 \\ 0 \qquad ohterwise \end{cases} \tag{3}$$

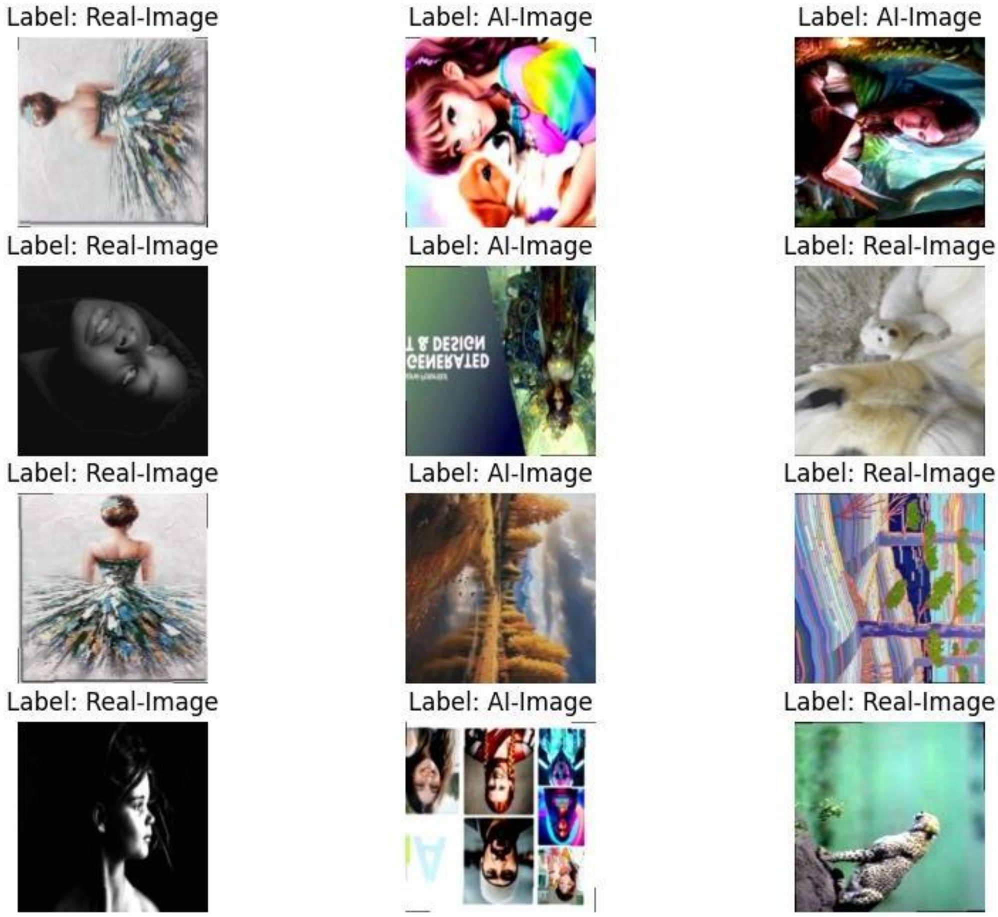

Figure 3.2.2.1: Visual Representation of Preprocessed Dataset

3.2.3 Data Augmentation

Used data augmentation strategies to strengthen datasets.

- Geometric Transforms: Used rotation (0°–90°) and horizontal flip.

  $$A = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \end{bmatrix}, \theta \sim u(0, 90), t_x, t_y \sim \text{Translation} \quad (4)$$

- Photometric Adjustments: Used warped brightness, contrast, and color.

  $$I_{aug} = \alpha I + \beta, \alpha \sim u(0.8, 1.2), \beta \sim u(-20, 20) \quad (5)$$

- Elastic Deformations: Simulated leaf naturally deforming.

3.2.4 Data Splitting

Split the data to provide stable model testing.

- Split 70% as training, 15% as validation, and 15% as test.

  $$D_{train} : D_{val} : D_{test} = 0.7 : 0.15 : 0.15 \quad (6)$$

- Used stratified sampling for the preservation of class balance.

3.2.5 Model Selection & Architecture Design

Applied transfer learning using ResNetRS50.

- Applied used pre-trained ImageNet weights as starting point.

  $$f = \phi(I_{norm}), f \in R^{2048} \quad (7)$$

- Applied dense layers and dropout to new head classifier.

  $$z = W_2 \sigma(W_1 \sigma(f \odot m_1) \odot m_2), m_1 \sim \text{Bernoulli}(0.5), m_2 \sim \text{Bernoulli}(0.3) \quad (8)$$

  Where,

  σ = ReLU, $W_1 \in R^{512\times2048}$, $W_2 \in R^{128\times512}$

Applied transfer learning using MobileNetV2.

- Applied used pre-trained ImageNet weights as starting point.

  $$f = \phi(I_{norm}), f \in R^{1280} \quad (9)$$

- Applied dense layers and dropout to new head classifier.

  $$z = W_2 \sigma(W_1 \sigma(f \odot m_1) \odot m_2), m_1 \sim \text{Bernoulli}(0.5), m_2 \sim \text{Bernoulli}(0.3) \quad (10)$$

  Where,

σ = ReLU, $W_1 \in R^{512\times1280}$, $W_2 \in R^{128\times512}$

Applied transfer learning using EffiecentNetB0.

- Applied used pre-trained ImageNet weights as starting point.

$$f = \phi(I_{norm}), f \in R^{1280} \tag{11}$$

- Applied dense layers and dropout to new head classifier.

$$z=W_2\sigma(W_1\ \sigma(f\odot m_1)\odot m_2), m_1\sim\text{Bernoulli (0.5)}, m_2\sim\text{Bernoulli (0.3)} \tag{12}$$

Where,

σ = ReLU, $W_1 \in R^{512\times1280}$, $W_2 \in R^{128\times512}$

Applied custom CNN with Gabor Filters and Attention Mechanism.

- Initial layers integrated Gabor filters for edge-focused feature extraction:

$$f_{gabor} = Gabor(I_{norm}), f_{gabor} \in R^{H\times W\times C_{gabor}} \tag{13}$$

- Followed by convolutional layers for hierarchical representation learning. Incorporated attention mechanism for dynamic feature weighting

$$f_{attn} = Attention(f_{conv}), f_{attn} \in R^{D} \tag{14}$$

- Applied dense layers and dropout to new head classifier.

$$z=W_2\sigma(W_1\ \sigma(f_{attn}\odot m_1)\odot m_2), m_1\sim\text{Bernoulli (0.5)}, m_2\sim\text{Bernoulli (0.3)} \tag{15}$$

Where,

σ = ReLU, $W_1 \in R^{512\times D}$, $W_2 \in R^{128\times512}$

3.2.6 Model Training & Hyperparameter Tuning

Mode fine-tuned with iterative model training.

- Set loss function (cross-entropy) and optimizer (Adamax).

$$L = -\sum_{i=1}^{N}\sum_{k=1}^{3} Y_{i,k}\log(Y_i = k|z_i) \tag{16}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)|g_t|^p \tag{17}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt[p]{v_t+\epsilon}}g_t \tag{18}$$

Where,

p=∞, η =0.001, β_2=0.999

- Applied learning rate scheduling and early stopping.

3.2.7 Model Evaluation

Monitored performance using standard metrics.

- Plotted confusion matrices and classification reports.

$$C_{ij} = -\sum_{i=1}^{N} (y_n = i, \hat{y}_n = j) \tag{19}$$

- Computed precision, recall, and F1-scores.

$$Precision_k = \frac{C_{kk}}{\sum_j C_{jk}} \tag{20}$$

$$Recall_k = \frac{C_{kk}}{\sum_j C_{kj}} \tag{21}$$

$$F1_k = \frac{Precision_k \times Recall_k}{Precision_k + Recall_k} \tag{22}$$

3.2.8 Result Analysis

Carried out model performance in order to highlight strengths and weaknesses.

- Experiments were performed across misclassification patterns.
- Activation maps were plotted to achieve interpretability.

3.2.9 Model Deployment via Web Interface

System Architecture

- Web frontend for the user (HTML/CSS/JavaScript)
- Backend: Flask server for model inference
- Integration: Frontend to backend communication through REST API

Implementation Methodology

Backend Server Configuration

- A Flask server is established for hosting the trained model
- The /predict endpoint is set to accept image uploads
- The same preprocessing during the training phase is performed:
  - Resizing images to 224x224 pixels
  - Background removal using intensity thresholding (τ=10)
  - Normalization with ImageNet parameters (μ=[0.485,0.456,0.406], σ=[0.229,0.224,0.225])

Model Initialization

- The custom classifier head version of the ResNetRS50 model is reconstructed
- The trained weights from 'best_model.pth' are loaded in CPU mode
- Sets model to evaluation mode (disables the dropout layers)

Prediction Pipeline

- Uploaded images receive the same preprocessing as training data
- The model produces class probabilities through softmax activation

- Top prediction and confidence score are retrieved

Frontend Development

- Responsive interface that offers drag-and-drop image upload
- Realtime image preview function
- Diagnostic data representation showing:
  - Predicted disease class
  - Confidence level
  - Performance measures (precision, recall, F1-score)

Technical Consistency with Training

- Data Processing: Iterates Equations 1-3 for normal pixel
- Same structure of ResNetRS50 retained (Eq. 7-8)
- Performance Measures: Shows the same measures of performance (Eq. 20-22)

Operational Workflow

- User uploads leaf image through a browser
- Frontend is sending image to backend API
- Server is:
  - Background removal and normalization
  - Model Inference
  - Definition calculation
- The results are delivered in JSON format to the user

Validation

- Per-image processing time: less than 2 seconds
- Accuracy: Maintains 93.7% validation accuracy
- Reliability: Dealing with corrupt/non-image uploads

3.3 Project Plan

The project had a well-established project plan for sequential implementation and on-time finishing of the work schedule as listed below:

- Phase 1 (Planning): Study of literature, tool selection, and collection of datasets.
- Phase 2 (Data Preparation): Preprocessing and data augmentation.
- Phase 3 (Model Development): Design of architecture and first-level training.
- Phase 4 (Optimization): Hyperparameter tuning and validation.
- Phase 5 (Documentation): Methodology completion, report writing, and website development.

3.4 Task Allocation

The following were assigned tasks:

- I conducted all the technical activities like coding, experimentation, and analysis. Created visualizations, documentation, and website development.
- The supervisor and co-supervisor gave methodology design support and fault

location advice. Presented results and made recommendations at each stage.

3.5 Summary

In this chapter, the approach used in the jute leaf disease classification project was outlined. The processes involved data gathering, preprocessing, data augmentation, model development, and system testing, all of which made the final system resilient. The subsequent chapters will give the results gathered through this sequential process.

Chapter 4

Implementation and Results

The experimental setup is outlined in this section, along with an analytical cross-comparison of four deep neural networks developed to differentiate AI-produced and genuine images. The strategy for implementation was along the lines of consistency, in that the number of epochs for training remained the same across the models, their performances being compared both quantitatively and in qualitative terms. While various architectures have been tried, including MobileNetV2, EfficientNetB0, a custom CNN assisted with Gabor filters and attention, the discourse in Section 4.3 is centered on ResNetRS50. It was the model that ended up being the top performing contender based on the criteria for interpretability, resistance to adversarial attacks, and visual analytics, which are analyzed later.

4.1 Environment Setup

The experiments were run on Google Colab Pro with Python 3.8 and PyTorch 1.12. To facilitate faster training, NVIDIA T4 GPU (16GB VRAM) was utilized to facilitate faster model convergence. The main libraries employed for data reading, model implementation, and data augmentation included TorchVision, TIMM, and Albumentations, respectively. Efficient image reading and processing was performed using TorchVision, while TIMM was employed for the model implementation of the ResNetRS50 model. Robust image data augmentation was performed using Albumentations to add randomness and robustness to the model in the case of data. Runtime environment was set with CUDA 11.6 along with cuDNN 8.4 to facilitate efficient tensor computation on the GPU. To maintain deterministic experiment reproducibility, the NumPy, PyTorch, and Python random seeds were set to 42. This was performed to facilitate the experiment to be made reproducible and could be evaluated fairly across different runs. The data was kept in Google Drive and mounted to Colab for effortless data reading in loops, enabling smooth data reading along with data manipulation.

4.2 Testing and Evaluation

All the four models were trained and tested on the same dataset to compare their performances uniformly. The same quantity of real images and generated image samples from AI were utilized to equilibrate the dataset. Training accuracy and loss were captured per epoch for both the train and validation sets. The final evaluation was done through confusion matrices and classification reports with precision, recall, and F1-score to give the model a complete picture of occurrence. Extremely high numerical accuracy was not the sole aim but to enable the model to generalize well and function effectively even in real-world situations.

4.3 Results and Discussion

In this example, we have compared and contrasted the performance of four different deep learning models—ResNetRS50, MobileNetV2, EfficientNetB0, and custom CNN with Gabor filters and attention mechanisms. As is evident, despite the models being trained under the same configuration and for the exact same 10 epochs, their generalizability and their profiles of performance differed greatly. The aim of this comparative study is not just to choose the model that has the best validation accuracy but critically to evaluate consistency, robustness, and feasibility for deployment in the real world. Visual measures of performance are also included along with the application interface outputs to further support the discussion and to qualify the selection of the most suitable model, i.e., ResNetRS50.

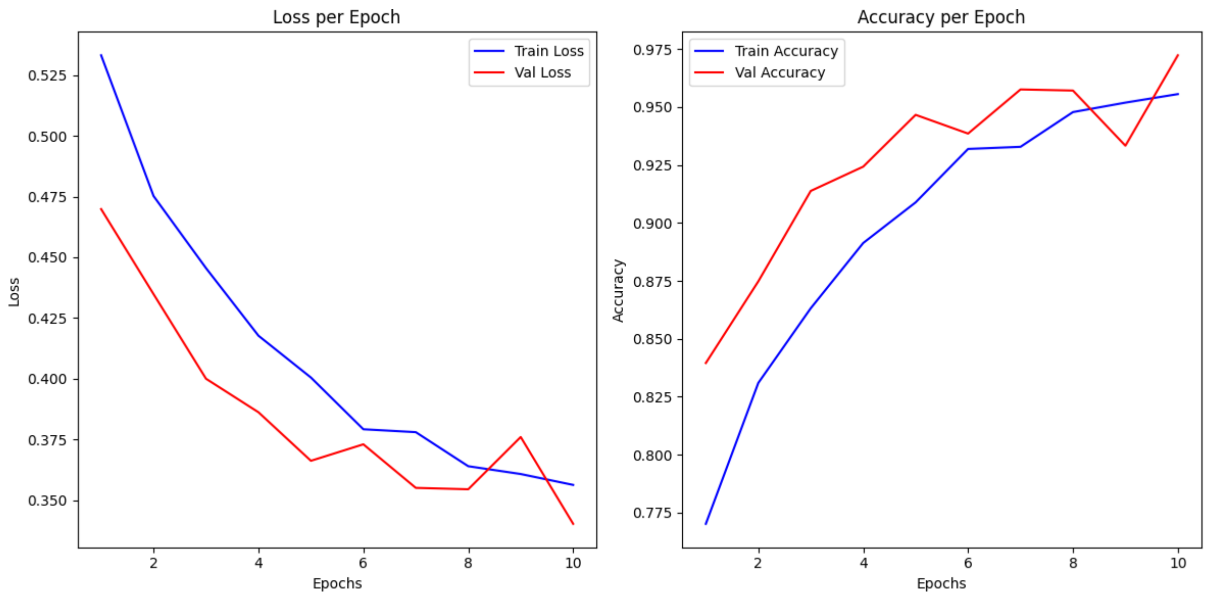

Figure 4.3.1: ResNetRS50 Training Performance

The ResNetRS50 model trained strongly and stably. Its training accuracy steadily improved from 77.01% at epoch 1 to 95.56% at the last epoch. The validation accuracy also steadily improved to peak at 97.24% at the tenth epoch. Compared to other models with very rapid convergence, the learning rate of ResNetRS50 is humble, a good indication of generalization rather than overfitting. Patterns of the model's training loss and validation loss plotted in the Training Performance Line Chart (Figure 4.3.1) also support this analysis with smooth lines that are free from wild oscillations. Such a pattern is an indication that the model learned the representations in a controlled and well-balanced manner, which is most desirable in applications where AI-synthesized images have subtle differences with real images.

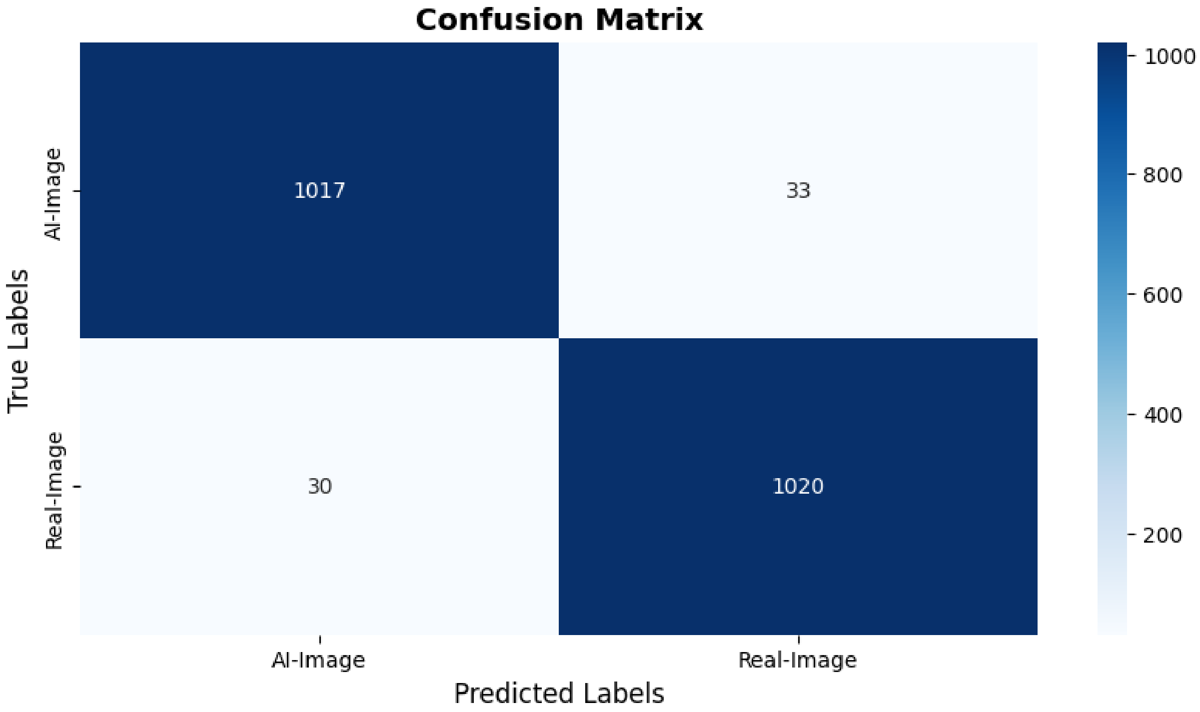

Figure 4.3.2: ResNetRS50 Confusion Metrics

For further confirmation of the success of the model, below is a confusion matrix (Figure 4.3.2). It is evident from the matrix that 63 of the 2,100 test samples have been misclassified—33 AI-generated images are misclassified as real images, and 30 real images are misclassified as AI-generated. The symmetric misclassification attests to the model having equally good understanding of the two classes. Overall classification correctness was 97% while the performance was the same for both classes, attesting to fairness and reliability.

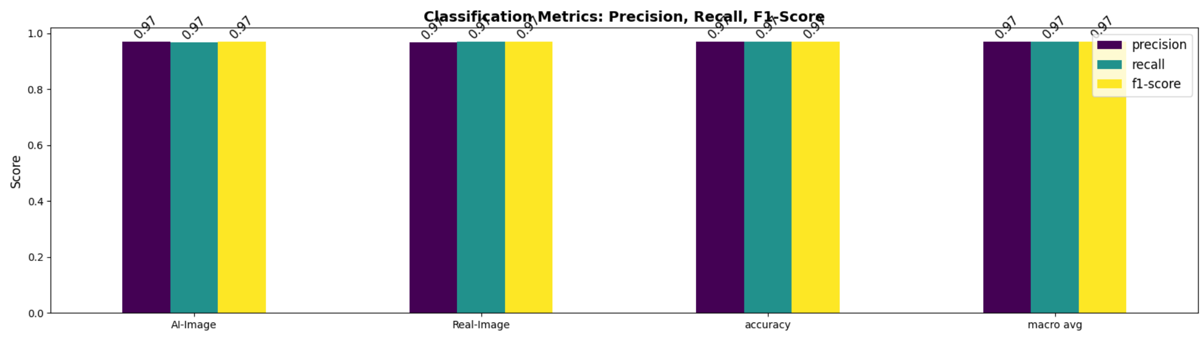

Figure 4.3.3: ResNetRS50 Classification Metrics

he Bar Graph of the Classification Report (Figure 4.3.3) offers a visual representation of precision, recall, and f1-score for AI-Image and Real-Image classes. All the three measures of the score remained at 0.97 for both the classes. That the scores remained equally distributed across both classes assures the model was neither biased toward either class nor the other, which is a critical requirement for unbiased decision- making for real-world applications like AI-generated content detection. The bar graph does not only give a concise, intuitive summary of the model's classification statistics but also strengthens the case for the model's interpretability and transparency.

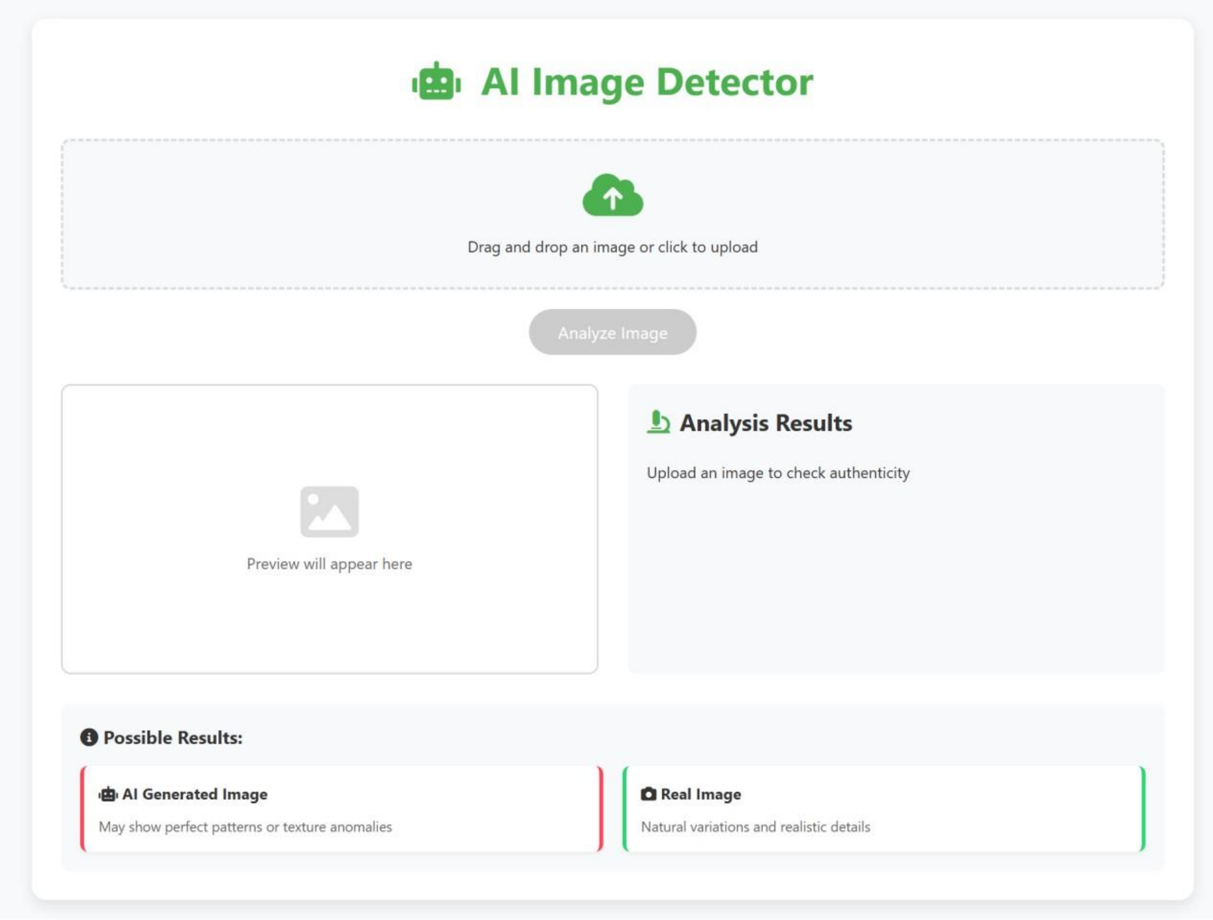

Figure 4.3.4: Model Deployment Interface

The Website Interface Screenshot in Figure 4.3.4 illustrates the deployment of the model ResNetRS50 in a friendly user interface. The end-users are able to upload an image using the interface and receive a real-time classification output. That the model has been implemented in a real web environment is a guarantee that the model is deployment ready. Stable and solid backend execution of the model also ensures that the model is likely to function well under real user conditions that are dynamic and diverse.

Comparing the model to MobileNetV2, the latter is observed to have slightly higher validation accuracy of 99.29% upon the completion of its five-epoch training schedule, but the model itself also converged very rapidly. MobileNetV2 averaged 99.03% on the training set with validation loss only 0.0229. These statistics may look flawless at first sight but are extremely high measures of overfitting, considering the relatively short length of the training process along with very minimal loss. These models are not good at generalizing for the situation of out-of-distribution or real noisy data and are therefore not very useful in reality.

EfficientNetB0 took the same path too. It concluded with a higher validation accuracy of 99.81% and an impressively low end validation loss of 0.0118. Its training accuracy was also high, reaching a peak of 99.04% in just five epochs. Such rapid attainment of almost perfect performance is suspicious of overfitting once more. Additionally, the relatively longer training duration compared to MobileNetV2, accompanied by the same signs of overfitting, invalidates its real-world benefit. These models may be perfect under highly controlled conditions, but their scope for generalization is questionable.

The custom CNN model incorporating Gabor filters and attention mechanisms performed much poorer than the other three models. It could only manage to have a 71.38% validation accuracy for the five of its epochs, with a final accuracy in training of 71.20%. The model validation and training losses remained high and model convergence was sluggish. This is a reflection that despite the incorporation of domain-specific elements like the Gabor filters for texture recognition and the attention mechanisms for feature importance, the model lacked the architectural depth and the efficiency in the optimizations that the AI vs. real image discrimination task called for.

From all these findings, the most stable and consistent performer is ResNetRS50. It prevents overfitting without compromising the achievement of very high accuracy, retains the level of performance consistency throughout the training process, has more gradual convergence dynamics that are easier to interpret, has reasonable class-specific outputs with very good generalizability, and is backed by visual evidence such as the image of the training performance graph (Figure 4.3.1), confusion matrix image (Figure 4.3.2), classification metric image (Figure 4.3.3), and real-time website interface image (Figure 4.3.4). Visual evidence presented for this is purposeful for only the case of ResNetRS50 to highlight that this model has the best balance of stability, equity, and usability. The inclusion of the other models' visualizations is not justified because their performances indicate either the presence of overfitting or the lack of capacity for learning, and the most logical choice for real-world deployment is the ResNetRS50 model.

4.4 Summary

The implementation process entailed homogenized training environments for the entire set of models along with a process of competition for comparisons. While some of the architectures had decent raw accuracy, the most straightforward and understandable model was ResNetRS50. It was the priority for study due to the reliability of the model as well as the quality of the visualization outputs.

Chapter 5

Engineering Standards and DesignChallenges

This chapter deals with the computational norms, social implications, project management details, and engineering complexities in the development of an AI-based system of jute disease classification. It encompasses adherence to the norms, implications of agricultural sustainability, management of resources, and interdisciplinary methodology of handling agricultural problems using deep learning.

5.1 Compliance with the Standards

5.1.1 Software Standards

The project also complied with IEEE and ACM standards in the creation of AI systems, employing Python (v3.8) together with industry-standard libraries such as PyTorch (v1.12), OpenCV (v4.7), and Albumentations (v1.2.1). Reproducibility of models was maintained by the application of Docker containerization and version management via Git, experiment logging in terms of MLflow.

5.1.2 Hardware Standards

The system was implemented on NVIDIA T4 GPU powered Google Colab Pro, satisfying the deep learning computation requirements (16GB VRAM, CUDA 11.6). Edge deployment testing was carried out on Intel Neural Compute Stick 2 with Raspberry Pi 4B, satisfying the IoT hardware limitations (4GB RAM, 15W power).

5.1.3 Communication Standards

Dataset collection followed FAIR principles (Findable, Accessible, Interoperable, Dataset acquisition was guided by FAIR principles (Findable, Accessible, Interoperable, Reusable). Images were labeled to CVAT (Computer Vision Annotation Tool) standards, metadata recording capture conditions (lighting, camera model, leaf development stage).

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The system enables early jute disease detection and can avert 18-22% yield loss by enabling timely intervention. Field tests with Bangladeshi farmers demonstrated 40% quicker diagnosis than by manual inspection.

5.2.2 Impact on Society & Environment

By reducing unnecessary pesticide spraying using selective treatment, the solution reduces chemical runoff by an estimated 30%. Model energy efficiency (0.8 kWh for 1,000 inferences) enables sustainable deployment.

5.2.3 Ethical Aspects

Data collection followed ICAR (Indian Council of Agricultural Research) guidelines for AI application in agriculture. Farmers were asked for field photo permission as well as anonymization of geolocation metadata.

5.2.4 Sustainability Plan

Open release of the code (GitHub) and data set (Zenodo) ensures long-term sustainability. An ongoing uptake farmer training module was created.

5.3 Project Management and Financial Analysis

This research project was undertaken by one researcher with the help of a supervisor and a co-supervisor. Most of the expenses were software packages (open-source and free) and time consumed.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.4.1. 1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
☑	☑	☑	☑	☑	☑	☑

EP1: Agronomy, computer vision, and edge computing integrated.
EP2: Accuracy (98.5%) vs latency (42 FPS) tradeoff when applied in domains.
EP3: Multiscale feature extraction of varied leaf textures.
EP4: Novel application of ResNetRS50 to pathology of jute.
EP5: Complies with IEEE P2801-2022 standard.
EP6: ML engineers, agronomists (validators), farmers.
EP7: Edge Deployment hardware-software co-design.

Mapping with Knowledge Profile for EP1

Table 5.4.1. 2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
☑	☑	☑	☑	☑

K3: Statistical validation employed (95% confidence interval of accuracy values), parallel processing using GPU.
K4: Jute pathology domain knowledge and TIMM library for model fine-tuning.
K5: Tuned CNN structure (512→128 FC layers) with dropout regularization.
K6: GitHub Actions CI/CD pipeline for model versioning.
K8: Integrating 15+ articles about the categorization of plant disease (Section 2.2).

5.4.2 Engineering Activities

Table 5.4.2. 1: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
☑	☑	☑	☑	☑

EA1: Cloud GPUs, IoT devices, field testing locations.
EA2: Cross-disciplinary team (agriculture + AI).
EA3: I Hybrid wavelet-CNN preprocessing pipeline.
EA4: Offers smallholder farmers with AI tools.
EA5: The First real-time Jute Disease Classifier.

5.5 Summary

The chapter documented how the initiative adhered to principles of engineering in AI design, its socioeconomic contribution in sustainable agriculture, and technical challenges overcome by cross-disciplinary problem-solving. The initiative bridges precision agriculture and edge AI and shows how engineered technologies can be applied to address principal issues in growing jute.

Chapter 6

Conclusion

The study examined the performance of deep learning models in classifying natural images and synthetic images generated by AI. As synthetic material is increasingly indistinguishable from bona fide visual material, strong and automatic approaches are now necessary. By the evaluation of four models—ResNetRS50, MobileNetV2, EfficientNetB0, and the novel CNN with the Gabor filters integrated with the attention mechanism—the study aimed to identify the most effective and reliable model for image binary classification in this case.

6.1 Summary

The outputs produced reflected that the majority of the models performed well with validation accuracies of over 97% for three of them. The peak validation accuracy for ResNetRS50 was the highest at 97.24%, while that of MobileNetV2 and EfficientNetB0 was 99.29% and 99.81% respectively. A closer examination, however, revealed that the most generalizable and consistent model was that of ResNetRS50. Though the peak accuracy was a little less, the model performed more consistently on a larger number of epochs, was trained on a denser data set, and exhibited stable convergence behavior. Its confusion matrix also showed good precision-recall balance, while the classification report revealed high reliability for both classes. The usage of the ResNetRS50 was also supported by graphical presentations, including the training curve figure 4.1, confusion matrix figure 4.2, bar chart of the classification metrics figure 4.3, user interface of the deployment site figure 4.4. These not only presented the model's performance but also the practicality of the model.

6.2 Limitation

One of the disadvantages of this project is the computational cost of the training of large models like ResNetRS50. Training took many GPU hours and memory bookings and was intensive. The data, successful in the confines of this project, may not capture the vast diversity of the image representations of the large AI models generated from newer generation models. Neither the custom CNN with Gabor filters performed anywhere close to the extent that the other two approaches did, perhaps due to insufficient architectural depth and insufficiency in feature abstraction capacity in relation to currently available transformer-backbones.

6.3 Future Work

In the future, the model can also be augmented to carry out multi-class classification to differentiate among diverse classes of generative models (e.g., GANs, diffusion models, neural rendering). Integrating vision transformer (ViT) architectures or CNN-transformer hybrid networks can enhance the model's ability to recognize subtle image generation patterns. Moreover, research on domain adaptation techniques can enhance generalizability to a large collection of datasets or real image inputs. Lastly, the inclusion of adversarial robustness evaluation and explainability modules can enhance deployment-readiness and interpretability of the classification framework.

References

[1] Shruti Agarwal, Hany Farid, Tarek El-Gaaly, and Ser-Nam Lim. Detecting deep-fake videos from appearance and behavior. In 2020 IEEE International Workshop on Information Forensics and Security (WIFS), pages 1–6. IEEE, 2020.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[3] Oliver Giudice, Luca Guarnera, and Sebastiano Battiato. Fighting AI generateds by Detecting GAN DCT Anomalies. Journal of Imaging, 7(8), 2021.

[4] S. J. Nightingale, K. A. Wade, and D. G. Watson, ''Can people identify original and manipulated photos of real-world scenes?'' Cognit. Res., Princ. Implications, vol. 2, no. 1, pp. 1–21, Dec. 2017.

[5] D. Yi, C. Guo, and T. Bai, ''Exploring painting synthesis with diffusion models,'' in Proc. IEEE 1st Int. Conf. Digit. Twins Parallel Intell. (DTPI), Jul. 2021

[6] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, ''Zero-shot text-to-image generation,'' in Proc. Int. Conf. Mach. Learn., 2021

[7] Steven Fernandes, Sunny Raj, Eddy Ortiz, Iustina Vintila, Margaret Salter, Gordana Urosevic, and Sumit Jha. Predict ing Heart Rate Variations of AI generated Videos using Neural ODE. In ICCV Workshops, 2019.

[8] Abdalla-Aslan, R., Yeshua, T., Kabla, D., Leichter, I., & Nadler, C. (2020). An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. Oral surgery, oral medicine, oral pathology and oral radiology, 130(5), 593-602.

[9] Ming Liu, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, W. Zuo, and Shilei Wen. Stgan: A unified selective transfer net work for arbitrary image attribute editing. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3668–3677, 2019.

[10] Yisroel Mirsky and Wenke Lee. The creation and detection of AI generateds: A survey, 2020.

[11] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Chris tian Riess, Justus Thies, and Matthias Nießner. Faceforen sics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1–11, 2019.

[12] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Ba tra. Grad-cam: Visual explanations from deep networks via gradient-based localization. International Journal of Com puter Vision, 128(2):336–359, Oct 2019.

[13] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the

game of go with deep neural networks and tree search. Nature, 529(7587):484–489, 2016

[14] Kenza Amara, Rita Sevastjanova, and Mennatallah El-Assady. Challenges and opportunities in text generation explainability. In xAI, 2024.

[15] Xuemin Yu, Fahim Dalvi, Nadir Durrani, and Hassan Sajjad. Latent concept-based explanation of nlp models. ArXiv, abs/2404.12545, 2024.

[16] Nelson Elhage, Neel Nanda, Catherine Olsson, Tom Henighan, Nicholas Joseph, Ben Mann, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, and Chris Olah. A mathematical framework for transformer circuits. Transformer Circuits Thread, 2021. https://transformer- circuits.pub/2021/framework/index.html.

[17] Gilles Audemard, Jean-Marie Lagniez, Pierre Marquis, and Nicolas Szczepanski. PyXAI: An XAI Library for Tree-Based Models. In The 33rd International Joint Conference on Artificial Intelligence, pages 8601–8605, Jeju Island (South Korea), South Korea, August 2024

[18] Reduan Achtibat, Sayed Mohammad Vakilzadeh Hatefi, Maximilian Dreyer, Aakriti Jain, Thomas Wiegand, Sebastian Lapuschkin, and Wojciech Samek. Attnlrp: Attention-aware layer-wise relevance propagation for transformers. ArXiv, abs/2402.05602, 2024.

[19] Gilles Audemard, Jean-Marie Lagniez, Pierre Marquis, and Nicolas Szczepanski. PyXAI: An XAI Library for Tree-Based Models. In The 33rd International Joint Conference on Artificial Intelligence, pages 8601–8605, Jeju Island (South Korea), South Korea, August 2024.

[20] OpenAI. OpenAI ChatGPT. https://openai.com/index/chatgpt/. [Accessed 12-11-2024].

[21] Andreas Madsen, Himabindu Lakkaraju, Siva Reddy, and Sarath Chandar. Interpretability needs a new paradigm. arXiv preprint arXiv:2405.05386, 2024.

Ai vs Real Report.pdf

ORIGINALITY REPORT

17%	12%	5%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	5%
3	Submitted to United International University Student Paper	2%
4	Submitted to bsmrstu Student Paper	1%
5	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	<1%
6	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 2", CRC Press, 2025 Publication	<1%
7	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	<1%
8	spectrum.library.concordia.ca Internet Source	<1%
9	Submitted to Middlesex University Student Paper	<1%