Image Classification: AI-Generated and Real Images

Group Members:

- Sanya Madan (2021561)
- Parisha Agrawal (2021270)
- Brinda Muralie (2021140)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Can you guess which image is real and which is AI-generated?

Motivation

- Rapid AI image generation blurs the line between real and fake, impacting media and legal contexts.
- Inability to differentiate can lead to fake news, reputation damage, and legal issues.

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Literature Review

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

CIFAKE: Advancing AI-Generated Image Recognition

- The paper addresses the challenge of distinguishing between real-life photographs and AI-generated images.
- A synthetic dataset mirroring the CIFAR-10 dataset is generated, providing a contrasting set of images for comparison.
- The study proposes using a Convolutional Neural Network (CNN) for binary classification of the images into ‘Real’ or 'Fake’.
- After training 36 network topologies, the optimal approach achieved a classification accuracy of 92.98%.
- The study implements explainable AI to identify features useful for classification, focusing on small visual imperfections in the image backgrounds.

GenImage: Advancing AI-Generated Fake Image Detection

- The paper introduces the GenImage dataset, a million-scale benchmark for detecting AI-generated images.
- The dataset includes over one million pairs of AI-generated fake images and real images.
- It covers a broad range of image classes and uses state-of-the-art generators, including advanced diffusion models and GANs.
- The paper proposes two tasks for evaluating detection methods: cross-generator image classification and degraded image classification.
- The GenImage dataset allows researchers to expedite the development and evaluation of superior AI-generated image detectors.

Generalizable Synthetic Image Detection via Language-guided Contrastive Learning

- The paper proposes a synthetic image detection method via language-guided contrastive learning.
- It augments training images with carefully-designed textual labels for joint image-text contrastive learning.
- The synthetic image detection is formulated as an identification problem, differing from traditional classification-based approaches.
- The proposed LanguAge-guided SynThEsis Detection (LASTED) model improves generalizability to unseen image generation models.
- LASTED delivers promising performance, exceeding state-of-the-art competitors by +22.66% accuracy and +15.24% AUC.

Dataset

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Dataset Description

- Dataset includes 120,000 images, evenly split between real and synthetic (fake) images, categorized into ten distinct classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck, with 6,000 images per class)

- The dataset is divided into 100,000 training images (50,000 each for real and fake images) and 20,000 testing images (10,000 each for real and fake images), all in RGB format and resized to 32x32 pixels.

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Dataset Visualisation

- TSNE plot (Figure 1) unsuitable due to scattered real and fake images.
- Pixel intensity histogram (Figure 2) displays differences between real and fake images.

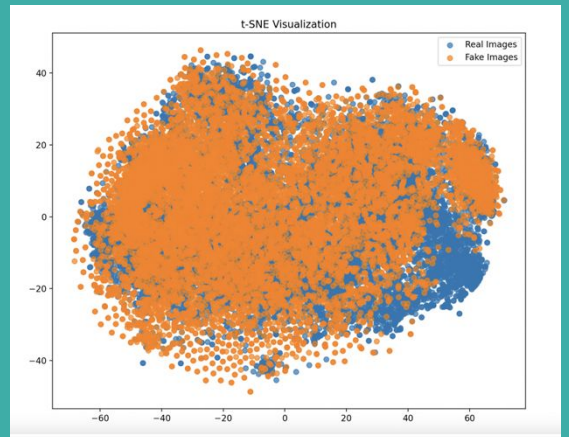


Figure 1. t-SNE Visualisation

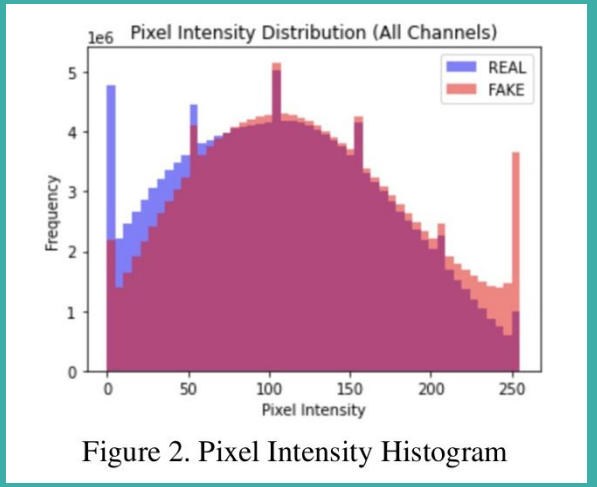


Figure 2. Pixel Intensity Histogram

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Dataset Visualisation

- Individual color intensity histograms reveal peak intensity disparities as potential distinguishing features for machine learning models.

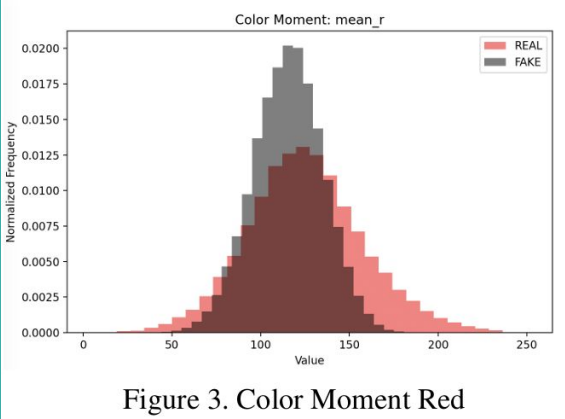


Figure 3. Color Moment Red

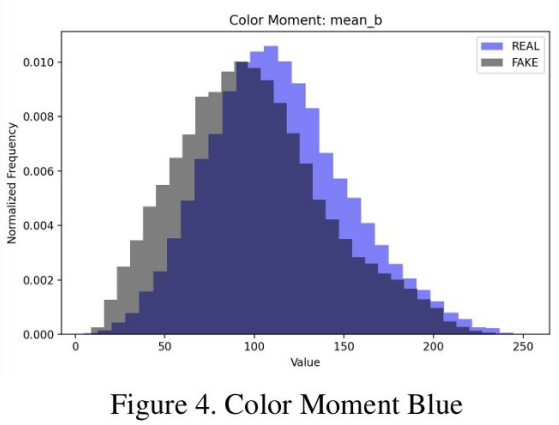


Figure 4. Color Moment Blue

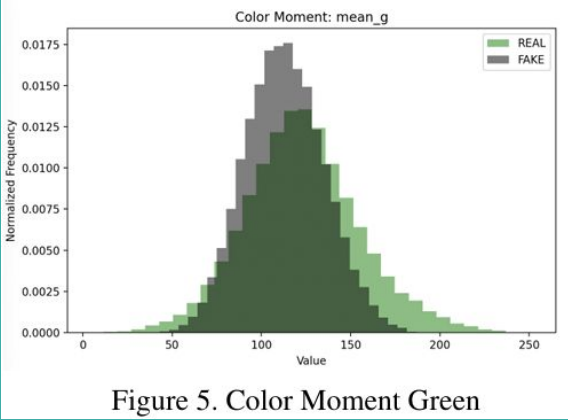


Figure 5. Color Moment Green

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Dataset Preprocessing

- Images loaded using CV2's 'imread' function and converted to numeric data with Numpy.
- Uniform 32x32 pixel dimensions ensured using OpenCV.
- Class labels transformed to numeric values using scikit-learn's LabelEncoder.
- No outliers found via data visualization and box plot (Figure 6).
- PCA deemed unnecessary as the dataset already had optimal dimensionality which was verified by experiments.

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

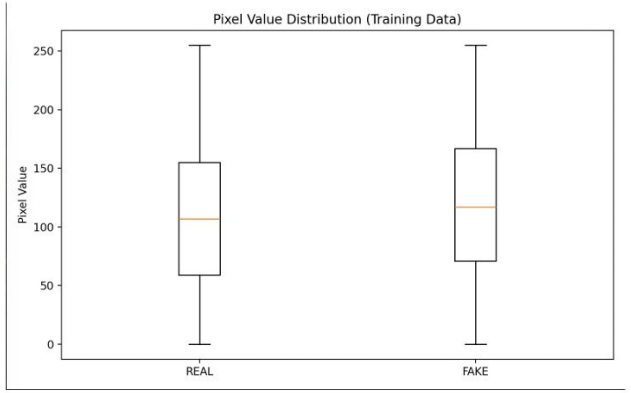


Figure 6. Box Plot

Number of components	Accuracy
10	0.71375
50	0.795
100	0.81175
500	0.81125
1000	0.81325
1024 (32 × 32)	0.8135

Table 1. PCA components and accuracy

Methodology

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

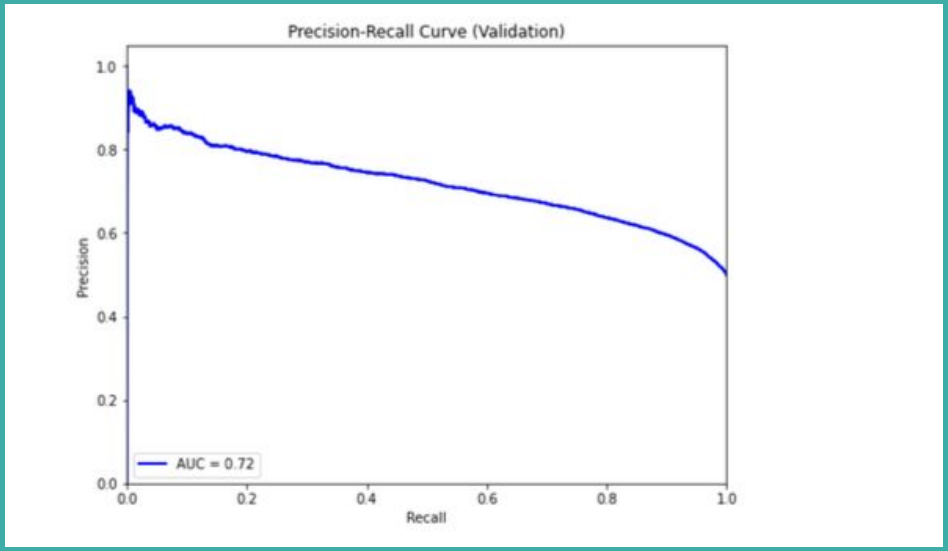
Methodology

- Employed Logistic Regression, Naïve Bayes classifier, Decision tree Classifier and Random Forest Classifier.
- We have used 4 models to train the dataset and evaluated accuracy scores for each model to compare them.
- Used Scikit learn, matplotlib, pytorch, and pandas library to implement this.
- Also plotted Precision recall curve and calculated various values to study each model.

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

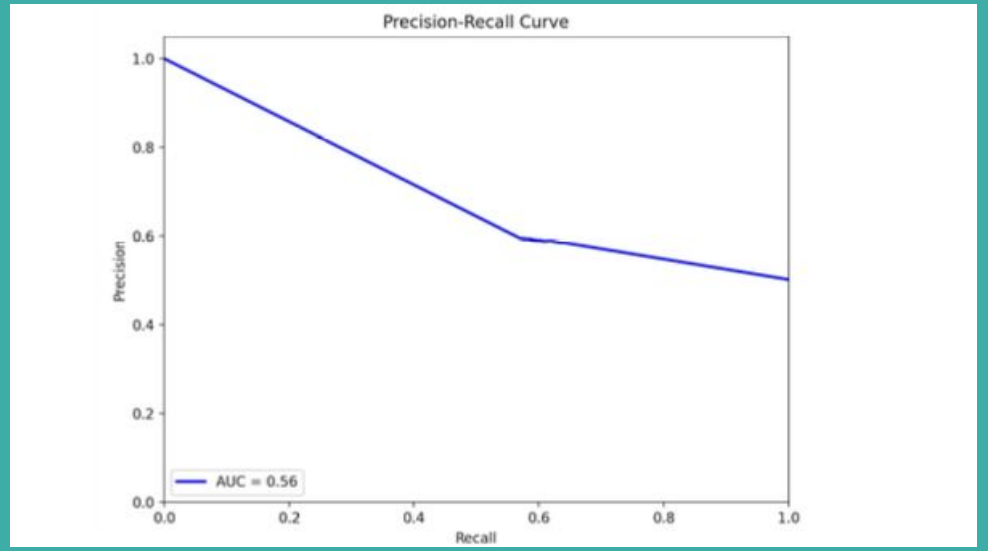
Logistic Regression

Value	Validation	Test
Accuracy	0.6793	0.67725
Precision	0.66386	0.6926
Recall	0.7218	0.6374
F1-Score	0.6916	0.6638
Specificity	0.6370	0.6374
Confusion Matrix:	[[6393 3642] [2772 7193]]	[[6374 3626] [2829 7171]]
False Positive Rate	0.3629	0.3626



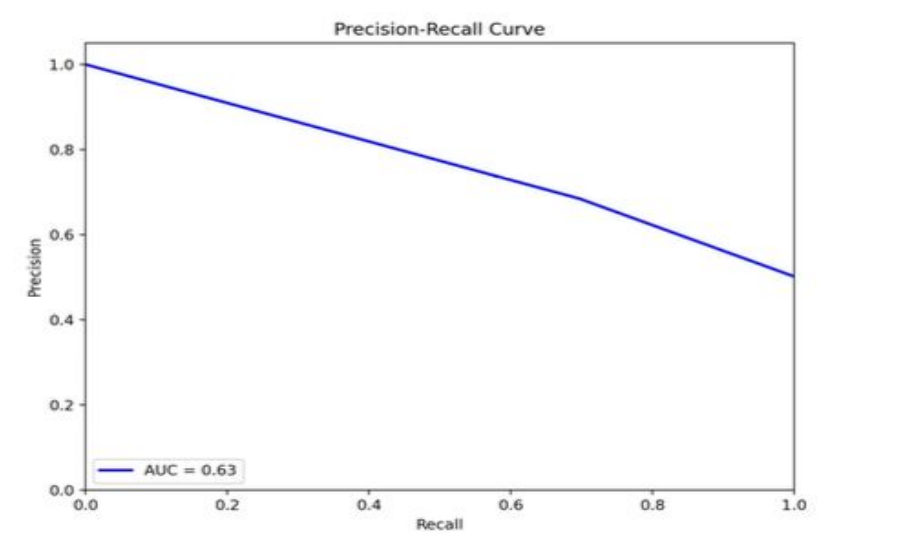
Naïve Bayes Classifier

Value	Validation	Test
Accuracy	0.5893	0.59275
Precision	0.59412	0.59570
Recall	0.57269	0.5773
F1-Score	0.58321	0.5863
Specificity	0.60602	0.6082
Confusion Matrix:	[[6039 3926] [4288 5747]]	[[6082 3918] [4227 5773]]
False Positive Rate	0.3939	0.3918



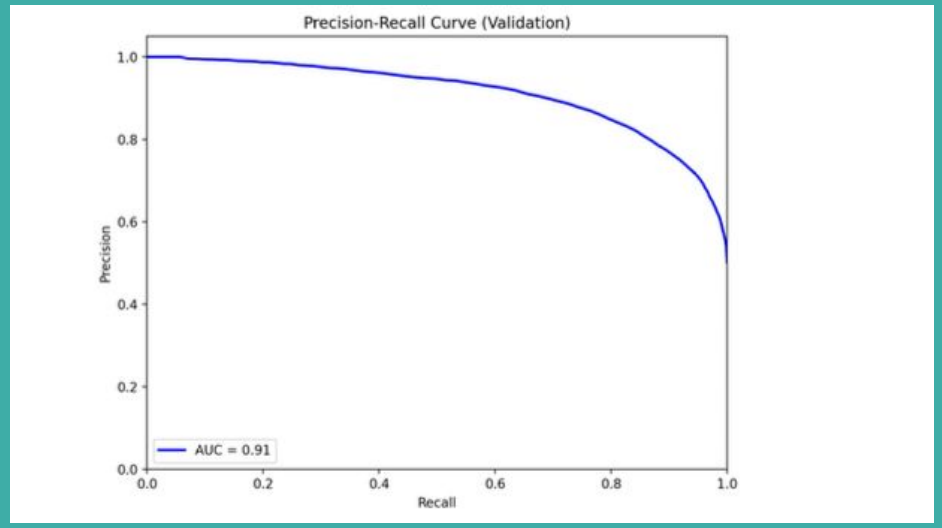
Decision Tree Classifier

Value	Validation	Test
Accuracy	0.6869	0.69745
Precision	0.68457	0.6983
Recall	0.6972	0.6953
F1-Score	0.6908	0.6967
Specificity	0.6764	0.6996
Confusion Matrix:	[[6741 3224] [3038 6997]]	[[6996 3004] [3047 6953]]
False Positive Rate	0.3235	0.3004



Random Forest Classifier

Value	Validation	Test
Accuracy	0.8272	0.82925
Precision	0.8498	0.80630
Recall	0.7963	0.8667
F1-Score	0.8222	0.8354
Specificity	0.8583	0.8667
Confusion Matrix:	[[8553 1412] [2044 7991]]	[[8667 1333] [2082 7918]]
False Positive Rate	0.1416	0.1333



Results and Analysis

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Results and Analysis

- We evaluated four machine learning models, and found varying levels of performance.
- Random Forests achieved the highest test accuracy of 83.07%, due to its ability to handle complex image data.
- Naïve Bayes exhibited the poorest performance with a test accuracy of 59.275%, due to its simplistic assumption of feature independence.
- Decision Trees performed reasonably well with a test accuracy of 69.765%, but are susceptible to overfitting.
- Logistic Regression showed moderate performance with a test accuracy of 67.725%. Its linear nature limits its capability to capture intricate patterns and relationships present in image datasets.

Model	Validation Accuracy	Test Accuracy
Logistic Regression	0.6793	0.67725
Naïve Bayes	0.5893	0.59275
Decision Tree	0.69345	0.69765
Random Forest	0.8266	0.8307

Timeline

- We were able to follow the proposed timeline and have researched and visualised the dataset, analysed various data preprocessing methods and trained and compared different ML models.
- For the further part of the project, we will be working on further optimisation and using SVM and Neural Networks (multilayer perceptron and CNN).

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Individual Contributions

- All team members have equal contribution in the project.
- Data Visualization - All 3 members
- Data Preprocessing - All 3 members contributed and analysed
- Model Training
  - Logistic Regression - Sanya
  - Naive Bayes - Brinda
  - Decision Tree, Random Forest - Parisha
- Report and Presentation Writing - All 3 members

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

References

1. Will Cukierski. (2013). CIFAR-10 - Object Recognition in Images. Kaggle. https://kaggle.com/competitions/cifar-10
2. Bird, J.J., Lotfi, A. (2023). CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. arXiv preprint arXiv:2303.14126.
3. M. Zhu, H. Chen, Q. Yan, X. Huang, G. Lin, W. Li, Z. Tu, H. Hu, J. Hu, and Y. Wang, “Genimage: A million-scale benchmark for detecting ai-generated image,” arXiv preprint arXiv:2306.08571, 2023.
4. Haiwei Wu, Jiantao Zhou, Shile Zhang, ”Generalizable Synthetic Image Detection via Language-guided Contrastive Learning”, arXiv preprint arXiv:2305.13800, 2023.
5. Krizhevsky, A. (2009) Learning Multiple Layers of Features from Tiny Images. Technical Report TR-2009, University of Toronto, Toronto. https://www.cs.toronto.edu/ kriz/learning-features-2009-TR.pdf

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Thank You!

Group Members:
- Sanya Madan (2021561)
- Parisha Agrawal (2021270)
- Brinda Muralie (2021140)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI