SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)
(Established under Section 3 of the UGC Act, 1956) | Re-accredited by NAAC with 'A++' grade | Awarded Category – I by UGC
॥वसुधैव कुटुम्बकम॥
50 Years of Excellence Since 1971

DeepFake Detection in Digital Forensics

Presented By:
ANJALI SINGH
RASHMI KADU
ROHINI BHARNE

Supervised By:
Dr. Priya Dasarwar

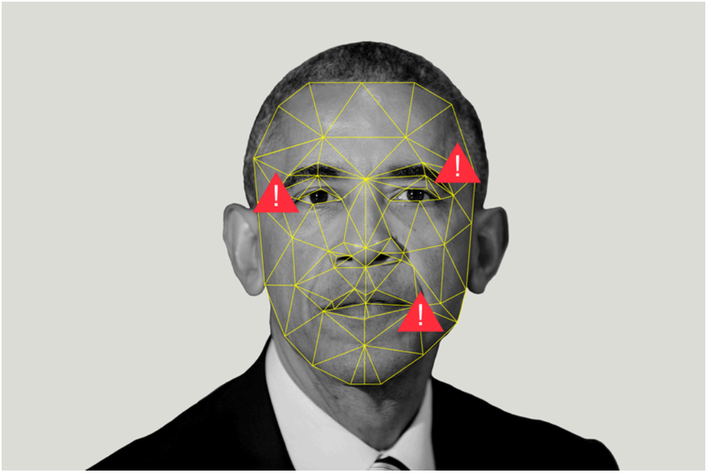
Symbiosis Institute of Technology, Nagpur Campus (SIT-N)

Introduction

- Rapid AI advancements have led to the rise of DeepFakes- realistic digital alterations that distort image, video and audio.
- These DeepFakes can create misleading representations, posing a serious challenge in the digital age.
- The project's goal is to detect and prevent the spread of manipulated media effectively.

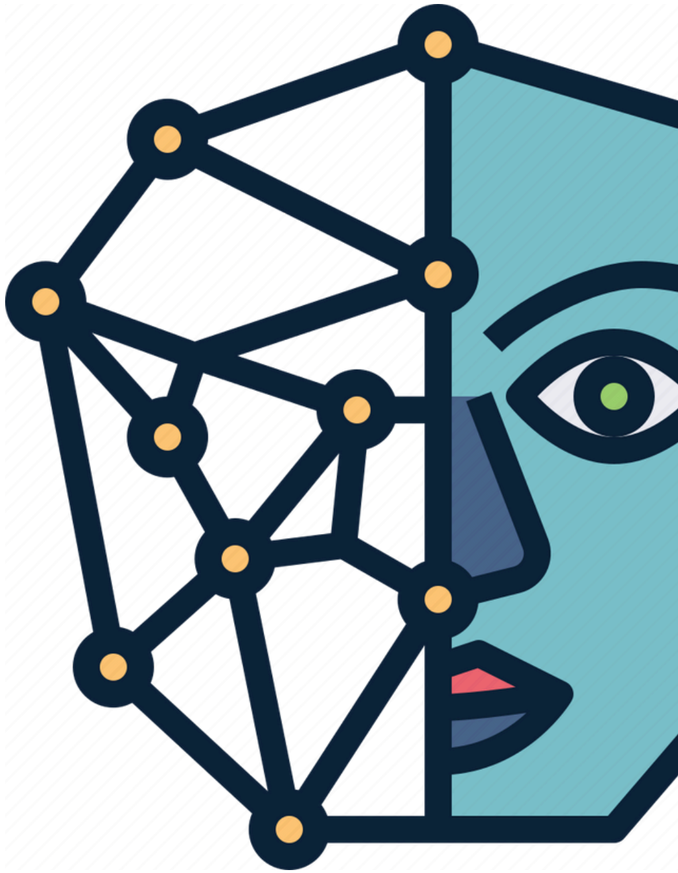
Background

- DeepFakes have become a critical issue in the digital world due to their increasing realism. The difficulty in distinguishing between genuine and manipulated content is a growing concern.
- The rise of DeepFakes is driven by sophisticated AI techniques, especially in machine learning. These techniques enable seamless alterations in audio and video files.
- This project aims to build on existing research and methodologies to address this issue.
- It will utilize Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) to analyze media for signs of tampering. The project contributes to the broader effort to combat misinformation.



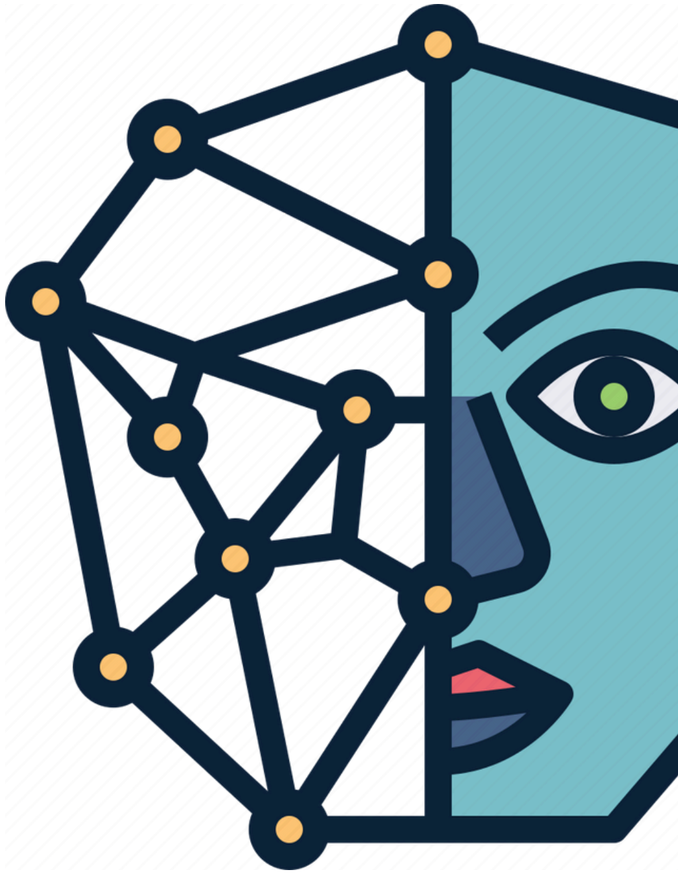
Problem Statement

- DeepFakes are becoming increasingly sophisticated, making them harder to detect with traditional methods.
- There is an urgent need for a reliable detection system capable of identifying manipulated media files with high accuracy.
- Developing an algorithm that can detect subtle discrepancies in media files, such as minor alterations in facial expressions, movements, or voice, remains a major challenge.
- Current detection techniques often struggle with real-time detection, especially for large-scale data, requiring improvements in computational efficiency.
- The detection system must be accessible and user-friendly, allowing non-experts to easily identify potential DeepFakes.
- The solution should also be adaptable to evolving manipulation techniques, as DeepFake generation methods rapidly improve.



Limitations of Previous Models

- Accuracy: Many previous models suffer from lower accuracy, especially when detecting high-quality or less obvious manipulations.
- Generalization: Older models often fail to generalize across different datasets or manipulation techniques, reducing their effectiveness in real-world applications.
- Computationally Intensive: Some detection methods require extensive computational power, making them impractical for large-scale or real-time detection.
- User Accessibility: Many existing models are designed for researchers or experts, with interfaces that are not intuitive for general users.
- Bias in Datasets: Several models are trained on biased or limited datasets, affecting their performance on diverse media files, especially when new manipulation techniques emerge.



Objective

1) Utilizing Comprehensive Datasets: Training the detection algorithm on a vast dataset of authentic and manipulated media to improve accuracy.

2) Developing a Detection System: Implementing deep learning models (CNNs, Xception, NN, etc) to identify anomalies in image contents.

3) Contributing to Misinformation Mitigation: Supporting efforts to combat digital misinformation by providing a tool that can identify DeepFakes.

4) Creating a User-Friendly Interface: Designing an interface that allows easy verification of media authenticity, even for non-technical users.

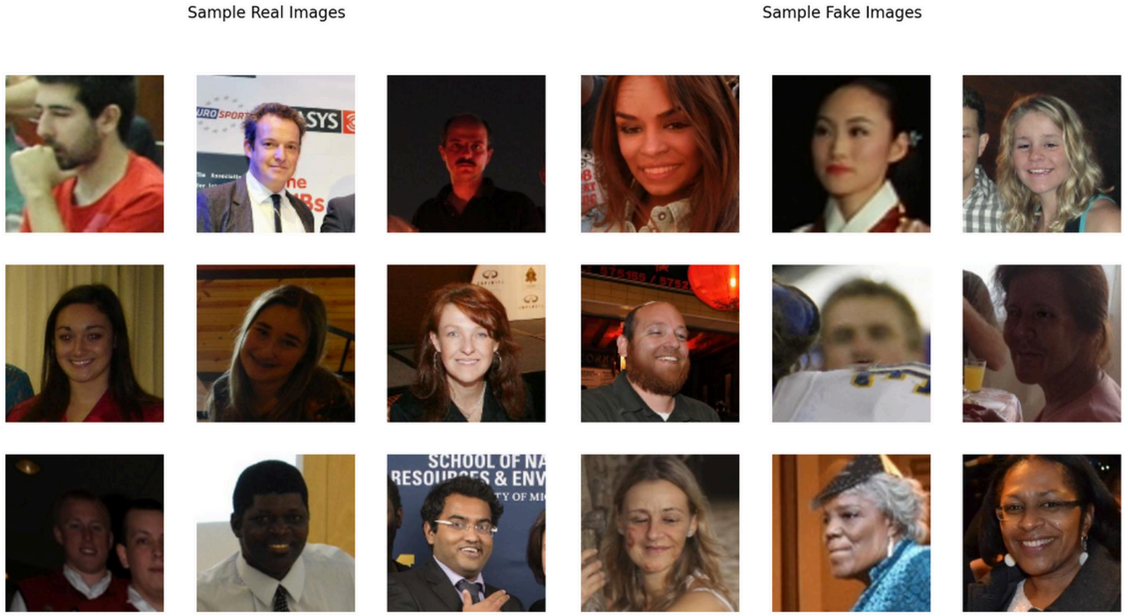
Literature Review

RefNo	Title	Methodology	Remarks
[1]	Samuel Henrique Silva et al. Deepake Forensics Analysis (2021)	Weakly supervised models, CNNs, explainable AI, DeeperForensics-1.0 dataset	Highlights need for transparency in AI detection systems.
[2]	Gaurav Gupta et al. Comprehensive Review of DeepFake Detection(2024)	Machine learning, ensemble methods, CNNs, feature fusion techniques	Focuses on generalization challenges and lightweight architectures.
[3]	Bahar Uddin Mahmud et al. Deep Insights of Deepfake Technology (2020)	Reviews deep learning, false content creation	Emphasizes both risks (disinformation) and benefits (film restoration).
[4]	Thanh Thi Nguyen et al. Deep Learning for Deepfakes Creation and Detection (2019)	Autoencoders, GANs, detection methods	Calls for solutions to counter deepfake impacts on society.
[5]	M. M. El-Gayar1 et al. Novel Deep Fake Detection Using GNNs (2024)	Graph Neural Networks (GNNs)	Demonstrates superior GNN performance for detecting deepfakes.
[6]	MD Shohel et al. Deep Face Detection Review (2022)	Traditional, machine learning, and deep learning approaches	Highlights ethical considerations in deepfake detection.
[7]	Amala Mary et al. Deep Fake Detection Using Deep Learning Techniques (2023)	CNNs, RNNs, transformers; analysis of facial expressions, motion, and audio-visual discrepancies	Focuses on real-time detection challenges and evolving attack methods.
[8]	Rimsha Rafque et al. Deep Fake Detection Using ELA and Deep Learning (2023)	Error-Level Analysis (ELA), CNNs, ResNet18, SVM, KNN	Achieves high accuracy (89.5%) using ResNet and KNN for detection.
[9]	Preeti et al. GAN-Based Deepfake Detection (2023)	GANs (DCGANs), Inception Score, FID	Discusses challenges of evolving GAN-based deepfake techniques.
[10]	Dafeng Gong et al. Deepfake Forensics Using CNNs (2020)	Custom CNN, VGG19, DenseNet-121, data augmentation	VGG19 achieves highest accuracy (95%); highlights importance for legal/security contexts.

DATASETS

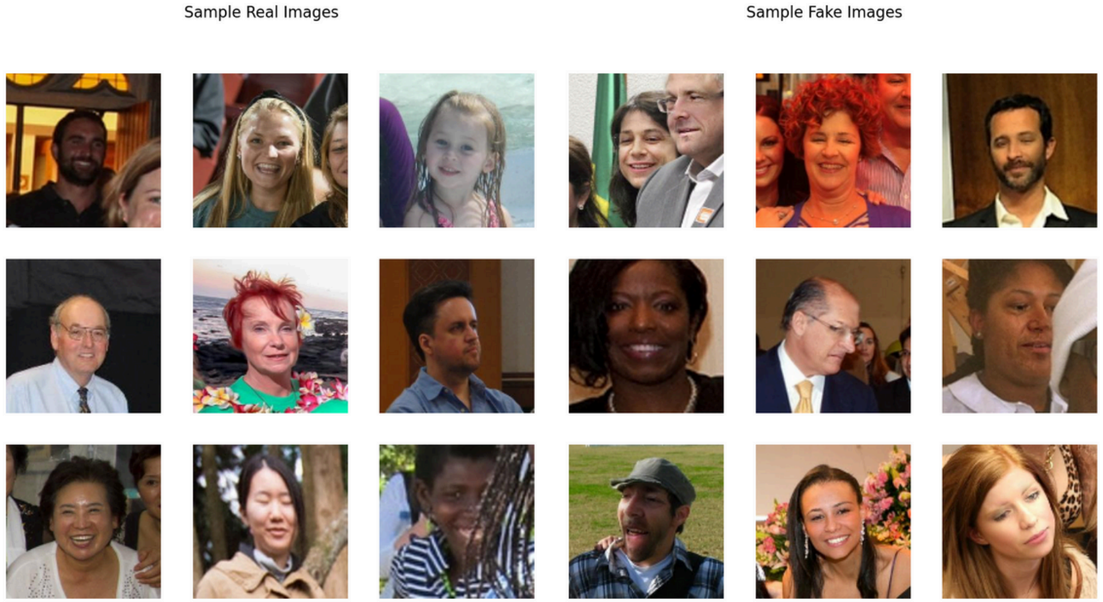
DATASET 1 (2 GB) :

The 'Deepfake and Real Images' dataset on Kaggle contains a collection of images labeled as either real or deepfake. It's designed for training and testing machine learning models to differentiate between authentic and manipulated (deepfake) images, providing a valuable resource for tasks like image classification and deepfake detection.

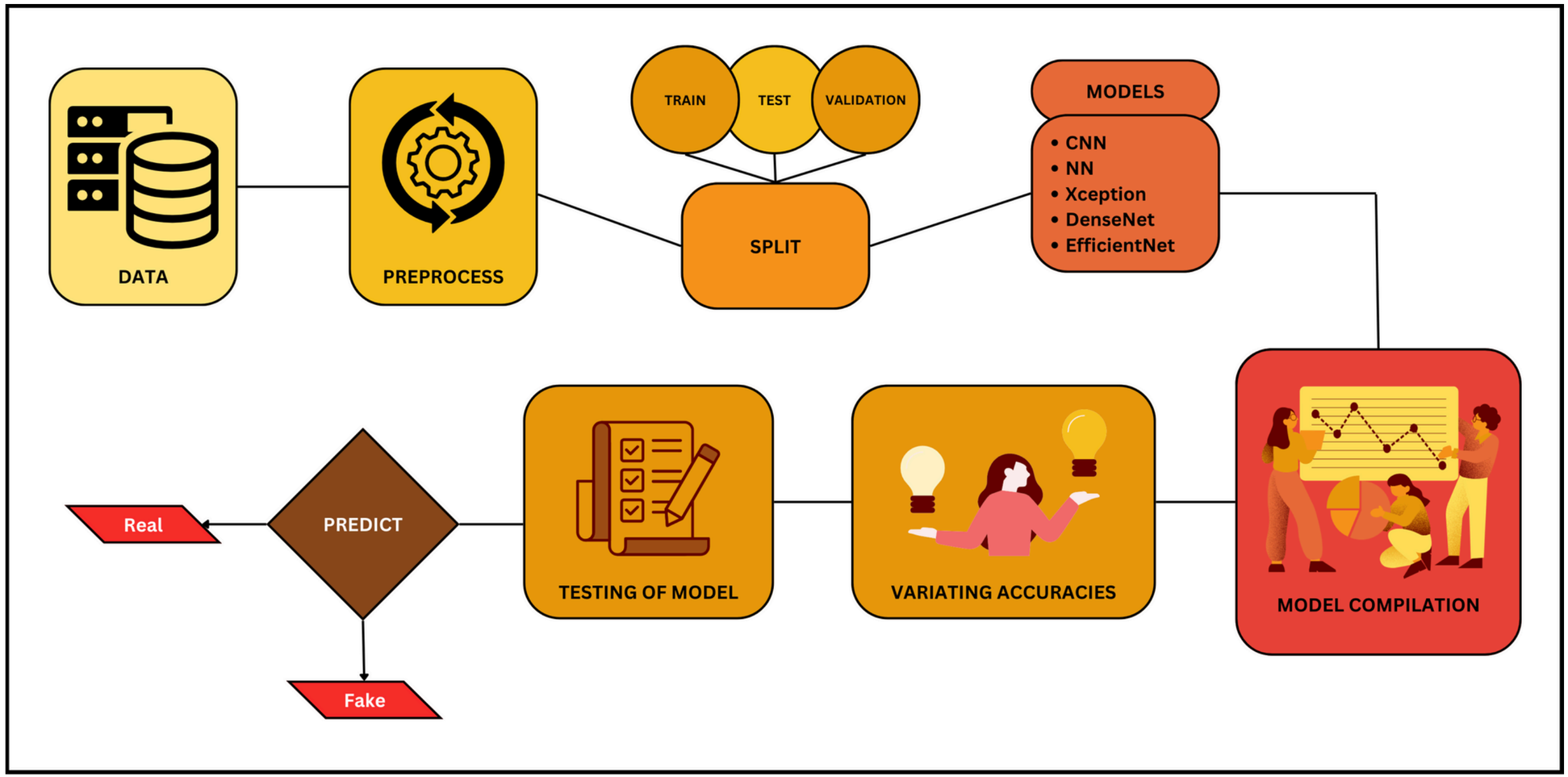


DATASET 2 (24 MB) :

The 'Bigger Dataset for Image Deepfake Detection' is a Kaggle dataset designed for training models to detect deepfake images. It includes a large collection of labeled images (real and fake), which can be used to train machine learning models for deepfake detection tasks.



System Architecture



About System Architecture

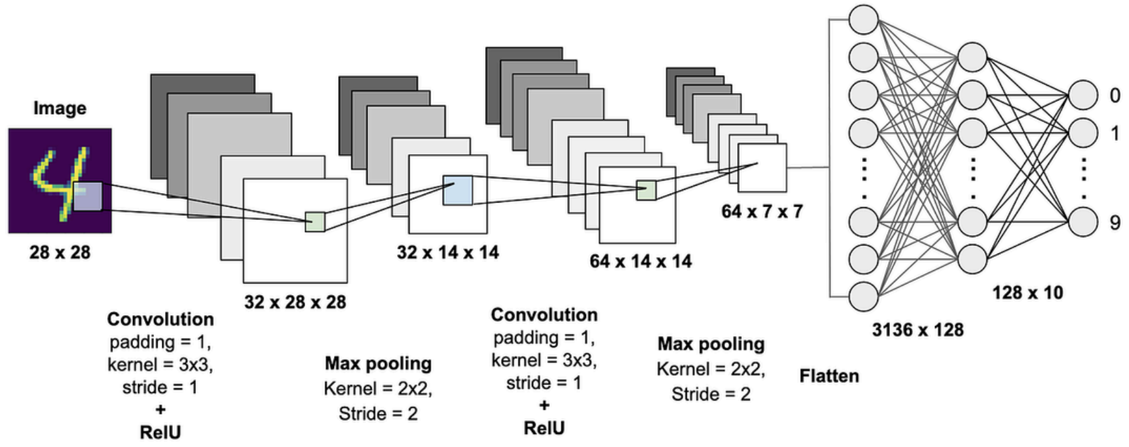
1. **Data Collection:** The system begins with the collection of data, typically images or videos. The data may consist of both real and fake (deepfake) content.
2. **Preprocessing:** Before analysis, the raw data is preprocessed. This involves operations like resizing, normalization, and augmentation, ensuring the data is in a suitable format for the model.
3. **Data Splitting:** The dataset is split into training, testing, and validation sets. The training set is used to train the model, while the test and validation sets evaluate its performance.
4. **Model Selection:** Various machine learning models are applied for deepfake detection, including:
    - CNN (Convolutional Neural Networks)
    - NN (Neural Networks)
    - Xception
    - DenseNet
    - EfficientNet
    - These models are chosen based on their capability to recognize patterns in image and video data, which is crucial for distinguishing deepfakes from real content.
5. **Model Compilation:** The selected models are compiled, and their performance is optimized using relevant loss functions and optimizers, which enable them to learn from the data during training.
6. **Testing the Model:** After training, the models are tested on unseen data to evaluate their accuracy and efficiency in classifying real versus fake content.
7. **Varying Accuracies:** Different models may yield varying accuracies. These accuracies are compared to determine the best-performing model for deepfake detection.
8. **Prediction:** Once the model is finalized and tested, it is used to predict whether a given piece of content (image/video) is real or fake (deepfake). The system outputs a binary classification of either "Real" or "Fake."

Deep Learning Models

1. **CNN (Convolutional Neural Networks):**

- CNNs are widely used for image and video classification tasks, including deepfake detection. They are composed of several layers: convolutional layers, pooling layers, and fully connected layers, designed to automatically detect spatial hierarchies of features (like edges, shapes, and textures) in images.

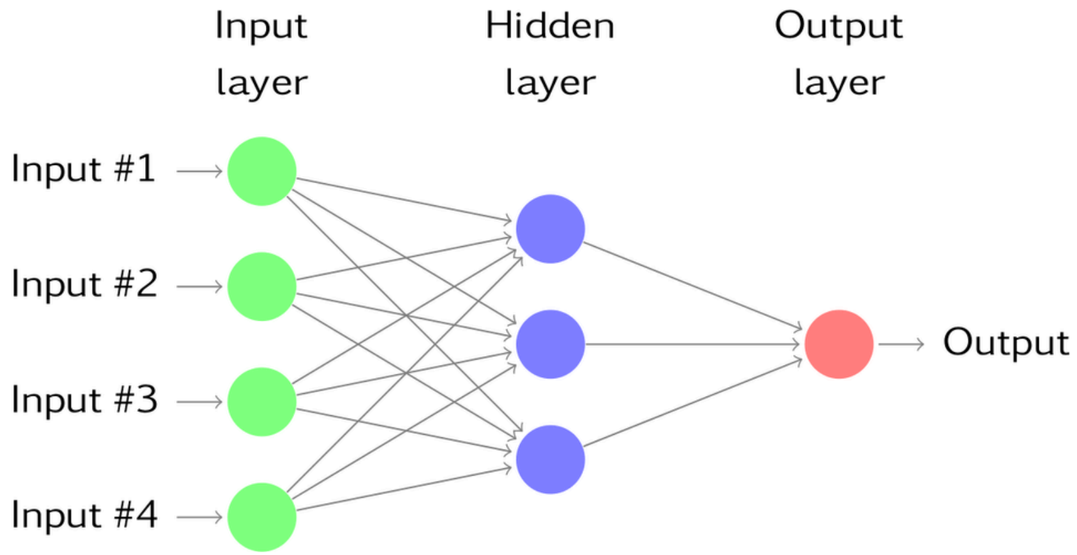
- Key Components:
- Convolutional Layers: Apply filters (or kernels) to the input data, extracting features such as edges, textures, or other important patterns.
- Pooling Layers: Downsample the feature maps, reducing their dimensions while retaining important features.
- Fully Connected Layers: Interpret the features and produce the final output



2. Neural Networks (NN) :

- Neural Networks (NNs), also known as feedforward networks, are a class of machine learning algorithms consisting of input layers, hidden layers, and output layers. Each neuron in one layer is connected to neurons in the subsequent layer, with weights assigned to these connections.

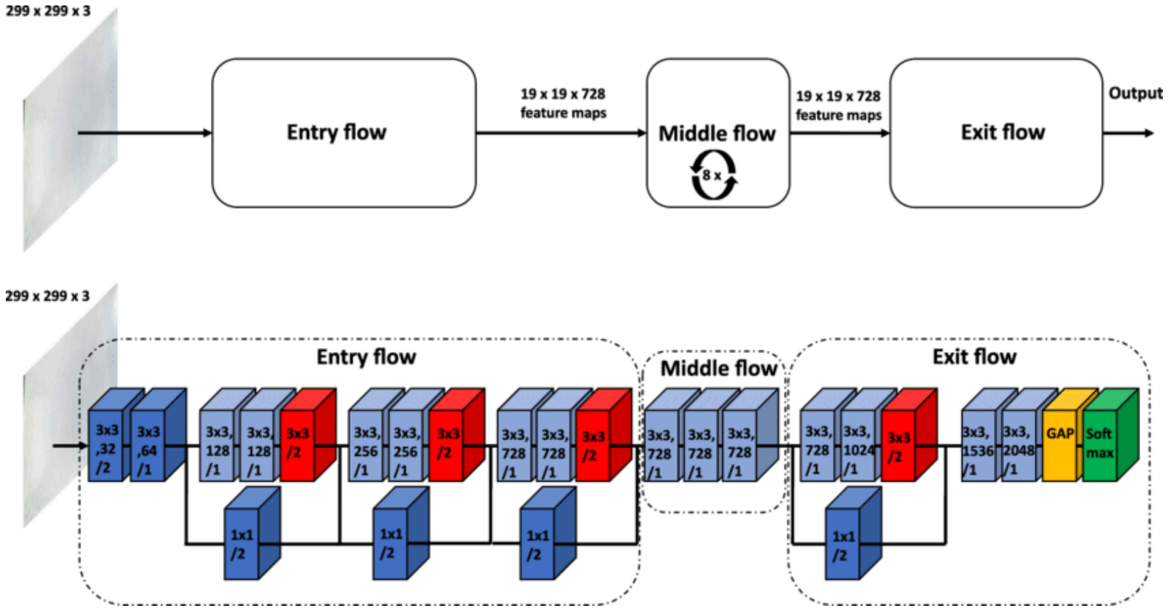
- Key Components:
- Input Layer: Takes in the data (image/video features in the case of deepfake detection).
- Hidden Layers: Multiple layers where data transformations occur, with each neuron applying a weighted sum followed by a non-linear activation function.
- Output Layer: Provides the final classification (Real or Fake).



3. Xception :

- Xception is an advanced deep learning architecture built as an extension of the Inception model. It replaces Inception modules with depthwise separable convolutions, leading to a more efficient and better-performing model for image classification tasks.

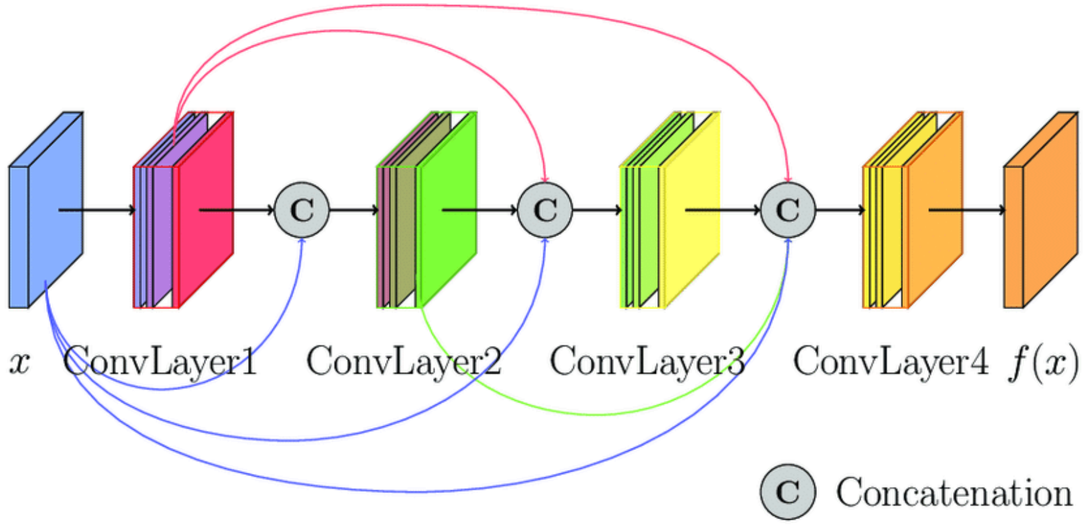
- Key Components:
- Depthwise Separable Convolutions: Factorize standard convolutions into two operations, reducing computational complexity while maintaining high accuracy.
- Residual Connections: Allow gradients to flow through the network efficiently, helping prevent vanishing gradient problems.



4. DenseNet :

- DenseNet (Densely Connected Convolutional Networks) connects each layer to every other layer in a feed-forward fashion. The key innovation of DenseNet is that instead of summing the inputs before passing to the next layer, the inputs are concatenated.

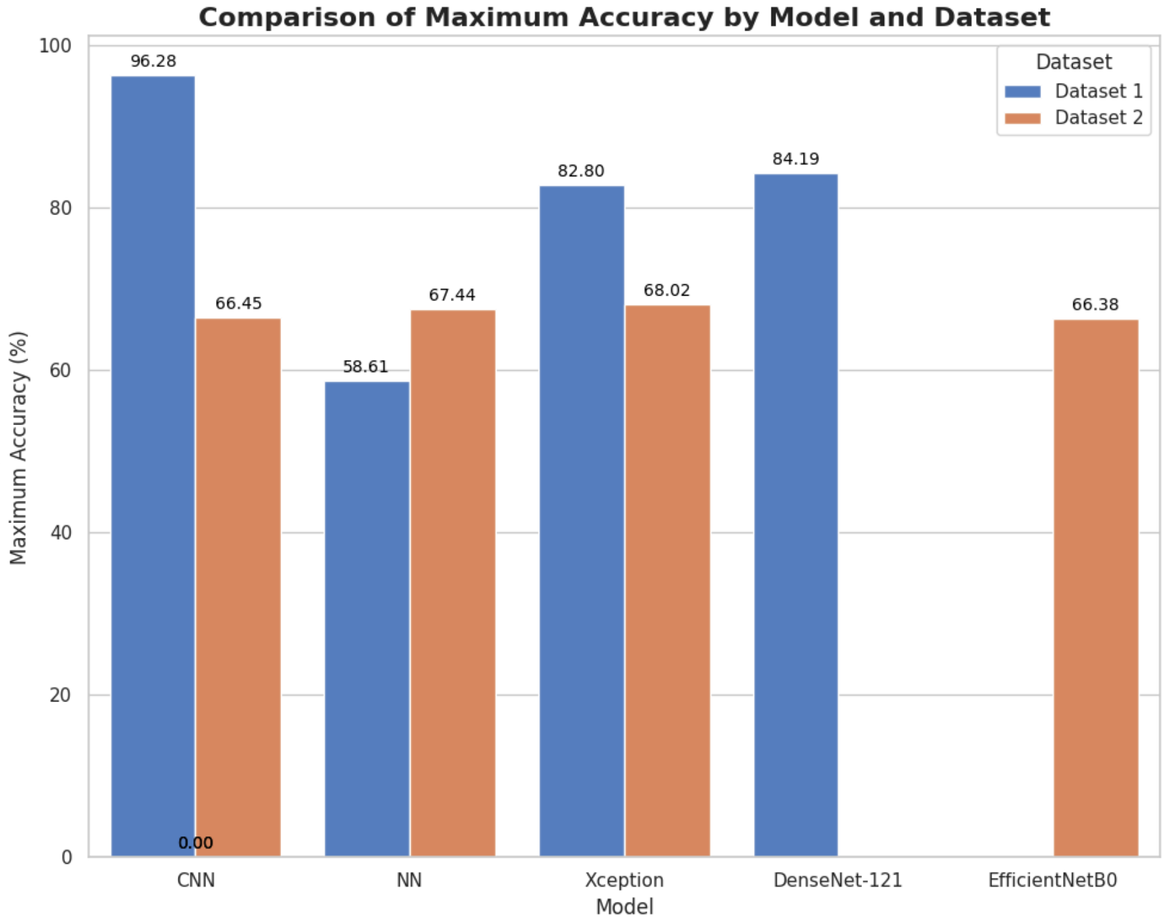
- Key Components:
- Dense Blocks: Each layer takes inputs from all preceding layers, which promotes feature reuse and reduces the number of parameters.
- Bottleneck Layers: Use 1x1 convolutions to reduce the dimensionality of feature maps, lowering computational cost.
- Transition Layers: Reduce the size of feature maps between dense blocks, helping in model compression.



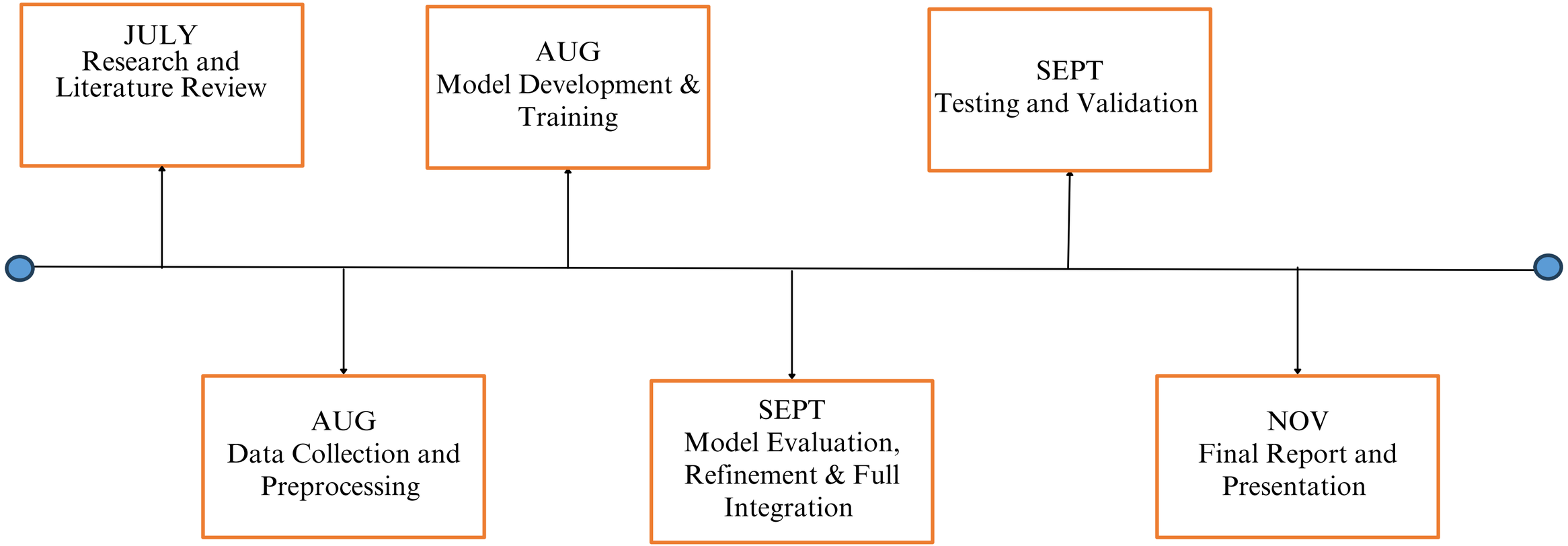
Observations and Analysis

TABLE I. MODELS AND ACCURACY SCORES

DeepFake Detection Models	Dataset 1	Dataset 2
CNN	96.28	66.45
NN	58.61	67.44
Xception	82.80	68.02
DenseNet-121	84.19	N/A
EfficientNetB0	N/A	66.38



Timeline


JULY
Research and Literature Review
AUG
Model Development & Training
SEPT
Testing and Validation
AUG
Data Collection and Preprocessing
SEPT
Model Evaluation, Refinement & Full Integration
NOV
Final Report and Presentation

Further Plans

- Working on learning more models for image classification to get great accuracy on collected datasets.
- Finding more appropriate dataset for image classification for building more efficient Model for deepfake detection.

References

[1] Silva, Samuel Henrique et al. “Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models.” Forensic Science International: Synergy 4 (2022): n. pag.
[2] Gupta, Gourav et al. “A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods.” Electronics (2023): n. pag.
[3] Mahmud, Bahar Uddin and Afsana Al Sharmin. “Deep Insights of Deepfake Technology : A Review.” ArXiv abs/2105.00192 (2021): n. pag.
[4] Nguyen, Thanh Thi et al. “Deep learning for deepfakes creation and detection: A survey.” Comput. Vis. Image Underst. 223 (2019): 103525.
[5] El-Gayar, M. M. et al. “A novel approach for detecting deep fake videos using graph neural network.” Journal of Big Data 11 (2024): 1-27.
[6] Rana, Md. Shohel et al. “Deepfake Detection: A Systematic Literature Review.” IEEE Access 10 (2022): 25494-25513.
[7] Mary, A. Victoria Anand and Anitha Edison. “Deep fake Detection using deep learning techniques: A Literature Review.” 2023 International Conference on Control, Communication and Computing (ICCC) (2023): 1-6.
[8] Rafique, Rimsha et al. “Deep fake detection and classification using error-level analysis and deep learning.” Scientific Reports 13 (2023): n. pag.
[9] Preeti et al. “A GAN-Based Model of Deepfake Detection in Social Media.” Procedia Computer Science (2023): n. pag.
[10] Gong, Dafeng. “Deepfake Forensics, an AI-synthesized Detection with Deep Convolutional Generative Adversarial Networks.” International Journal of Advanced Trends in Computer Science and Engineering (2020): n. pag.

Thank You...