Deepfake Detection using Deep Learning with InceptionV3

K. Cheritha, S. Akhil, V. Bhanu Prakash and A. Akhil Reddy

Dept Of Computer Science and Engineering, Malla Reddy University, Hyderabad

Mrs. Aurchana. P, Assistant Professor.

Dept Of Computer Science and Engineering, Malla Reddy University, Hyderabad

Abstract - Deepfake technology has rapidly evolved, making it increasingly difficult to distinguish between real and manipulated videos. This poses serious risks, including misinformation, identity theft, and digital forgery. To address this challenge, we propose a deep learning-based deepfake detection model that leverages InceptionResNetV2, a hybrid architecture combining the strengths of Inception networks and Residual networks (ResNet). Our approach efficiently extracts key facial features from video frames and classifies them as real or fake. The detection pipeline includes video preprocessing, frame extraction, feature extraction using InceptionResNetV2, and classification through a deep learning model. Our approach utilizes RGB colour space for feature extraction, ensuring that fine-grained visual artifacts present in manipulated videos are effectively captured. The detection pipeline involves video preprocessing, frame extraction, RGB-based feature extraction using InceptionResNetV2, and classification using a deep learning model. We train and evaluate our model using benchmark deepfake datasets, including Face Forensics++, Deep Fake Detection Challenge Dataset, and DeeperForensics-1.0, which contain diverse real and fake video samples. The model predicts whether a video is real or fake while providing a confidence score for better interpretability Experimental results demonstrate that our InceptionResNetV2-based model achieves high performance in deepfake detection. Our model achieves an accuracy of 81.3%, a precision of 82.84%, a recall of 84.00%, and an F1-score of 83.42.1%, indicating its effectiveness in distinguishing between real and fake videos. Future enhancements include real-time detection capabilities, adversarial training for improved robustness, and explainable AI techniques to provide greater transparency in deepfake detection. This research contributes to ensuring the authenticity of digital content and strengthening defences against deepfake-based cyber threats

Keywords – DFDC Deepfake Detection Challenge, CNN-Convolutional Neural Networks

Introduction

Deepfake technology refers to using artificial intelligence (AI), specifically deep learning techniques, to create highly realistic synthetic images, videos, and audio that can manipulate reality. These deepfakes are generated using generative adversarial networks (GANs) and autoencoders, making them difficult to detect with the naked eye. While deepfake technology has positive applications in entertainment and creative industries, it also poses serious threats, including the spread of misinformation, identity theft, and fraud.

The increasing sophistication of deepfake models necessitates the development of robust detection techniques. Traditional detection methods based on handcrafted features and statistical analysis are no longer sufficient due to the advanced nature of deepfake manipulations. As a result, deep learning-based approaches, particularly CNNs and architectures like InceptionV3 and ResNet, have gained prominence for detecting fake media. These models leverage powerful feature extraction mechanisms to identify inconsistencies, such as unnatural facial expressions, lighting anomalies, and artifacts that deepfake algorithms struggle to replicate perfectly.

The project aims to develop an efficient deepfake detection system using deep learning techniques. By leveraging pre-trained architectures like InceptionV3 and ResNet, the system can automatically extract crucial features from images and videos to differentiate between real and fake content. The implementation involves training these models on a large-scale dataset namely DFDC dataset+, optimizing them for high accuracy, and ensuring their robustness against adversarial deepfake attacks.

Literature review

In [1], The study by Aurchana et al. (2020) investigates the use of pre-trained convolutional neural networks (CNNs) for OSCC analysis at different stages. The study emphasizes the significance of early and precise OSCC detection with the help of deep learning algorithms, which can greatly enhance diagnosis and treatment planning. Through the utilization of pre-trained CNN models, the researchers hope to classify OSCC stages accurately, minimizing the use of manual histopathological examination. The article explains the benefits of employing CNNs in medical imaging, such as feature extraction, pattern recognition, and automatic classification, which improve diagnostic accuracy. The findings of the study highlight the potential of deep learning in medical imaging, opening the door to more efficient and accurate cancer detection systems.

In [2], Aurchana et al. (2024) introduce a research paper on ensemble-based soil classification via machine learning, highlighting enhanced accuracy in the identification of soil types. The study investigates different machine learning models and combines ensemble techniques to improve predictive accuracy. Through the integration of several classifiers, the research hopes to increase more accurate and credible soil classification, which is essential for agriculture and environmental research. The results prove the efficiency of ensemble learning in managing soil heterogeneity and enhancing classification results, indicating its possible use in smart agriculture and land management.

In [3], The study by Aurchana et al. (2023) aims to apply machine learning methods for leaf disease detection in the development of precision agriculture. The paper presumably makes use of multiple machine learning models, such as Convolutional Neural Networks (CNNs) for the identification of diseases in images and Support Vector Machines (SVM) or Random Forest for feature-based classification. CNNs are typically employed in plant disease diagnosis because they can learn deep features from images, whereas SVM and ensemble techniques assist in enhancing classification performance. The article emphasizes the importance of automatic disease detection in minimizing crop losses and maximizing agricultural productivity. Please let me know if you would like more information regarding the models utilized.

In [4], Aurchana et al. (2022) have suggested a mouth gesture classification through computational intelligence that improves human-computer interaction. The research potentially uses deep neural networks like Convolutional Neural Networks (CNNs) to extract features and classify, or machine learning-based methods like Support Vector Machines (SVM) for gesture detection. These identify mouth movements for enhancing accessibility apps, speech disabled communication, and interactive systems. The study emphasizes the effectiveness of computational intelligence in proper mouth gesture recognition, which contributes to the growth of assistive technology and gesture-based interfaces.

In [5] the authors provide a detailed analysis of Deepfake detection methods, emphasizing the superiority of deep learning approaches over traditional techniques. With the deep analysis of 112 papers the authors found out different detection models and algorithms being used to detect if a video/image/audio is real or fake, some of which are Deep learning models, particularly CNNs like XceptionNet, VGG, and ResNet & Machine learning techniques, including SVM, Logistic Regression, and Random Forest, rely on handcrafted features for classification, while statistical models use divergence measures to differentiate real and manipulated media.

The publication [6] introduces the (DFDC) dataset, the largest open-source face-swapped video dataset, containing more than100,000 clips from 3,426 paid actors. As the previous and other datasets limits in the size and diversity, the DFDC is one

of the best among them. The DFDC dataset has remarkably made a history through a Kaggle competition providing a benchmark for evaluating the diverse detection models where techniques like MTCNN for face detection and EfficientNet, Xception, and ensemble-based models proved highly effective. The study also speaks about the frame-based and video-based networks, and preeminent incorporation of 3D CNNs.

The authors of [7] discuss how different deepfake systems create fake images, videos and audio by replacing scenes or images, altering movies and modifying sounds in a way that make them in differentiable from the real content. The creation involves multiple deep learning (DL) algorithms to generate highly realistic synthetic media. It is evident from the review that deepfakes are generated using facial landmark extraction, autoencoders, CNN-based models, and adversarial techniques like face swapping, face reenactment, modifying audio, visual streams, attribute manipulation and many more.

According to the findings in [8], deep learning-based models outperform other approaches in identifying manipulated media, when trained on large datasets. The authors mainly worked on identifying the face swaps through patch-based detection framework. Also learnt personalized face action patterns like lip motions/mouth movements using the pre-trained lipreading network to fine tune the detection.

The Study [9] introduces a new framework called Diffusion Learning of Inconsistency Pattern (DIP) which uses a transformer-based model to identify inconsistencies. The framework starts with the spatiotemporal encoder, which extracts features from videos. To improve accuracy the authors decided to apply the Spatio Temporal Invariant Loss (STI Loss).

The review conducted in [10] offers a comprehensive evaluation of Deepfake classification emphasizing for accurate identification integrating with pre-trained convolutional neural networks mainly InceptionV3, XceptionNet, InceptionResV2, DenseNet-169, MobileNet, EfficientNet and NasNet-Mobile and ResNet-101. The study defines that Support Vector Machine (SVM) classifier combined with the DenseNet-169 provided the most accurate results as 98%. And second highest accuracy being acheived by XceptionNet 97.2%.

According to [11], Deepfake detection has evolved through multiple strategies, including deep learning, machine learning, and statistical models, with CNN-based models proving most effective. In this study the Universal Adversarial Perturbations, making attacks more accessible and feasible for widespread use. They have created a threat model making Deepfakes videos to evade detection, causing fake videos to be misclassified as real and vice versa. They have used $L\infty$ norm constraint for the perturbations.

According to the findings in [12], They explored the high-level architectures Convolutional Neural Networks (CNN) and Multilayer Perceptron (MLP) and observed that the implementation of InceptionV3 and InceptionResNetV2 models integrated with Multilayer Perceptron (MLP) as their high-level architecture achieved an impressive accuracy rate of 84.6%. And exceptionally the MLP InceptionResNetV2 shows impressive performance with 12 True Negatives (TN) and only 1 False Negative (FN).

The authors of [13] have employed three models, i.e. CNN-RNN (Inceptionv3-RNN) and two 3D CNNs(I3D and MC3) for deepfake classification in this study. The study shows that Q-learning algorithm has defeated the existing limitations like random or threshold-based selection with diverse swarm behaviors findings.

The study in [14] highlights the growing threat of Deepfake generation and explores various detection methods, comparing

their performance on benchmark datasets. Multiple pre-trained CNN networks were trained using tranfer learning and fine-tuned with different hyperparameters including optimizer, learning rate, batch size, and number of epochs. They implemented using six different configurations to test the effectiveness.

1. Framework of the proposed work:

The deepfake detection model leverages InceptionV3, a powerful CNN, along with a custom CNN model for robust classification between real and fake media. The dataset used for training and evaluation is sourced from the DFDC, which contains a diverse collection of both authentic and manipulated video frames. This dataset is crucial in training the model to recognize subtle visual inconsistencies introduced by deepfake generation techniques.

The InceptionV3 model is a state-of-the-art deep learning architecture that is pre-trained on large-scale image datasets like MobileNet. It is known for its efficient feature extraction capabilities using factorized convolutions, auxiliary classifiers, and batch normalization. By leveraging this pre-trained model, the deepfake detection system can quickly identify manipulated regions based on patterns learned from vast amounts of real-world images. The model is fine-tuned on the DFDC dataset to improve its ability to differentiate between authentic and synthetic content.

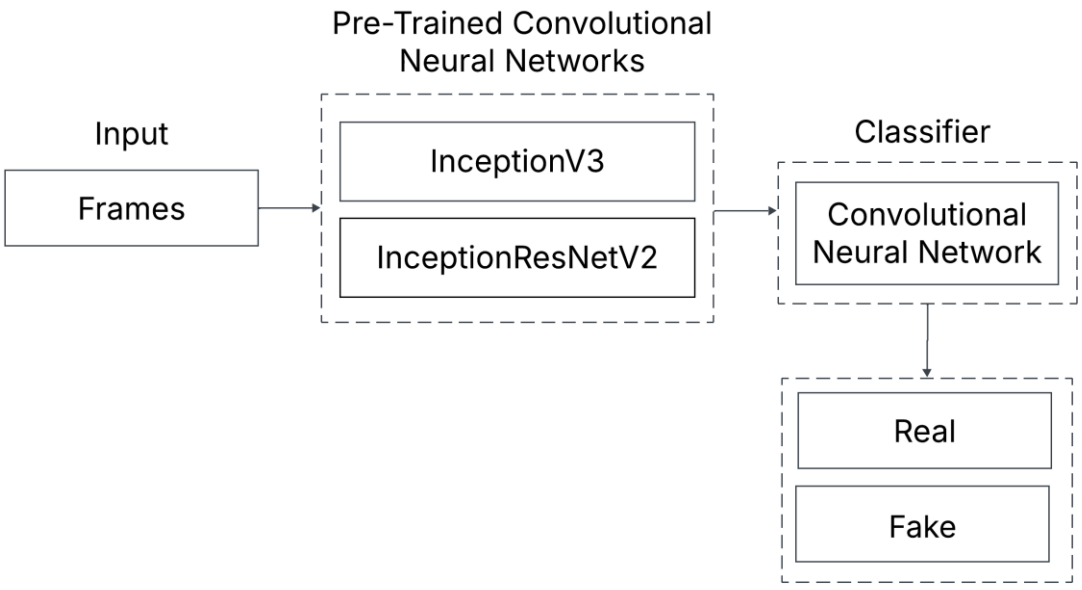

Fig 1: Overall Block Diagram of Proposed Work

Alongside InceptionV3, a custom CNN model is also designed to further refine feature extraction. This custom model consists of multiple convolutional layers with ReLU activations and batch normalization, progressively learn hierarchical representations of facial features. The architecture includes fully connected layers with dropout layers to prevent overfitting and enhance generalization. The final dense layer utilizes a sigmoid activation function to classify whether a given input is real or fake.

To optimize training, the system uses binary cross-entropy as the loss function and Adam optimizer for efficient weight updates. Performance is evaluated using accuracy, precision, recall, and F1-score, ensuring that the model is both effective and reliable. The combination of InceptionV3's transfer learning and a custom CNN model enhances deepfake detection, making it a robust solution for combating AI-generated media manipulations.

InceptionV3: Inception-v3 is a CNN structure that improves on previous versions of Inception by streamlining the framework and deploying additional inception modules than Inception-v2. It uses Factorized 7 x 7 convolutions, Label Smoothing, and an extra auxiliary classifier to transport label information lower down the network, among many other improvements (also uses batch normalization for layers in the side head). Suratkar et al. (2020) used different architectures such as Inception v3, MobileNet, ResNet50, etc. in task of detecting deepfakes using transfer learning. The Inception v3 model, according to their findings, outperformed all other models in terms of accuracy and predictions while requiring less training time and computational complexity. The Adam optimizer, with a learning rate of 0.0001, was found to be the most suitable choice for this system in terms of achieving the optimal solution in a short amount of time with high precision.

2. Preprocessing:

The videos of the collected dataset are pre-processed by splitting the videos into frames using cv2 library of python. The faces are cropped from the frames created and saved in a folder. We used pre-trained convolutional neural networks for the feature extraction in the preprocessing process. We trained over the dataset with 5 models namely, InceptionV3, ResNet50, Xception, InceptionResNetV2. Comparatively between all the models we observed that InceptionV3 and InceptionResNetV2 achieved highest accuracy exceptionally InceptionV2 has an efficient performance in feature extraction.

2.1 Feature Extraction

2.1.1 InceptionV3 is a deep learning model developed by Google that builds upon the original Inception architecture to enhance image classification performance. It introduces several improvements, including factorized convolutions, asymmetric convolutions, and optimized grid size reductions, making the network more efficient in terms of computational cost. The model processes images at a resolution of 299x299 and consists of approximately 23.9 million parameters. One of its notable features is the inclusion of auxiliary classifiers, which help improve gradient flow during training, leading to better convergence. InceptionV3 has been widely used in various computer vision tasks due to its balance between accuracy and efficiency. With a top-5 error rate of 3.46% on the ImageNet dataset, it is a preferred choice for applications like object detection, medical imaging, and transfer learning, where high accuracy is essential while keeping computational demands reasonable.

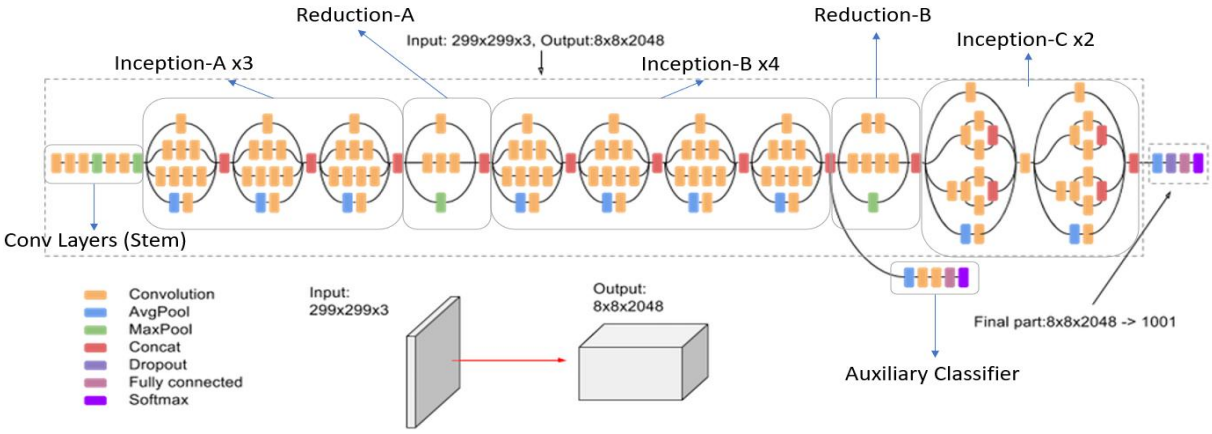

Fig. 2 Architecture of Inception-V3 Model

Stem (Initial Convolutional Layers):

The network begins with a stem module, consisting of convolutional layers (Conv layers) and max pooling operations. These layers extract basic low-level features such as edges, textures, and color variations from the input image (299×299×3). The stem efficiently reduces spatial dimensions while retaining critical feature representations.

Inception Modules (Feature Extraction):

InceptionV3 employs three types of Inception blocks (A, B, and C) for multi-scale feature extraction:

- Inception-A (×3) captures fine-grained details using small kernel convolutions.
- Reduction-A downsamples feature maps while maintaining essential spatial information.
- Inception-B (×4) expands the receptive field, learning mid-level patterns such as shapes and textures.
- Reduction-B further reduces dimensions, optimizing computation for deeper layers.
- Inception-C (×2) enhances abstract feature representation with deeper and wider convolutional layers.

Final Processing and Classification

After feature extraction, the network applies Global Average Pooling (GAP), replacing fully connected layers to reduce overfitting. A fully connected layer refines the learned representations, followed by a softmax output layer for final classification. In deepfake detection, this structure allows the model to detect subtle inconsistencies in manipulated videos by identifying anomalies in texture, lighting, and facial landmarks across multiple feature scales.

2.1.2 InceptionResNetV2 is an advanced convolutional neural network that combines the inception modules with residual connections to achieve improved training speed and accuracy. This hybrid architecture enables deeper networks without the common problem of vanishing gradients, allowing the model to learn more complex features effectively. Like InceptionV3, it also accepts 299x299 input images but has around 55.8 million parameters, making it a larger and more powerful model. The integration of residual connections reduces the risk of degradation in deeper networks, helping the model converge faster while maintaining high accuracy. With a top-5 error rate of 3.08% on the ImageNet benchmark, InceptionResNetV2 outperforms its predecessor, making it highly suitable for applications requiring detailed image analysis, such as autonomous systems, medical diagnostics, and advanced object recognition. Although it demands more computational resources, its ability to balance speed and precision makes it a top choice for deep learning-based vision tasks.

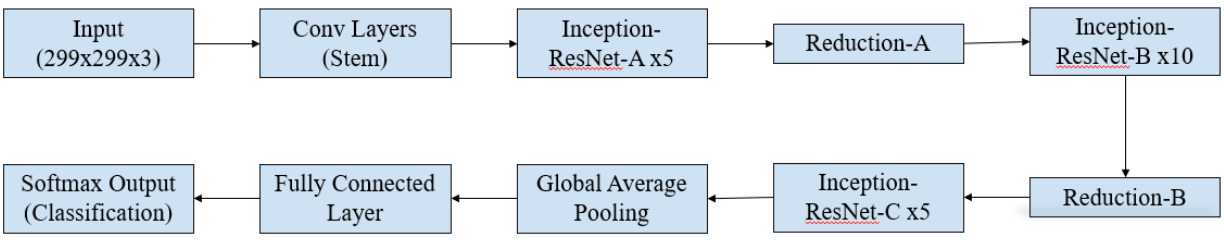

Fig. 3 Architecture of Inception-ResNet-V2 Model

maintaining high accuracy. With a top-5 error rate of 3.08% on the ImageNet benchmark, InceptionResNetV2 outperforms its predecessor, making it highly suitable for applications requiring detailed image analysis, such as autonomous systems, medical diagnostics, and advanced object recognition. Although it demands more computational resources, its ability to balance speed and precision makes it a top choice for deep learning-based vision tasks.

Table 1: Feature Comparision between InceptionV3 and InceptionResNetV2

Feature	**InceptionV3**	**InceptionResNetV2**
Input Size	299x299	299x299
Accuracy	~81.2%	~79.9%
Computational Cost	Moderate	Higher compare to InceptionV3
Scalability	Fixed	Flexible (B0 to B7)
Best Use Case	Deepfake Detection	General image classification
Parameters	~23M	~55M

2.2 Models- Convolutional Neural Network

In our model, we leveraged CNNs, particularly using InceptionV3 and InceptionResNetV2. These pre-trained architectures efficiently process image frames from videos, extracting meaningful features while maintaining computational efficiency.

Our model consists of several key layers that contribute to the feature extraction and classification process:

2.2.1 Convolutional Layers: The model consists of multiple convolutional layers with varying kernel sizes (3×3 and 5×5) to detect low-level features (edges, textures) in early layers and high-level abstract features (facial structures, deepfake artifacts) in deeper layers.

2.2.2 Batch Normalization Layers: Applied after convolutional operations to stabilize and accelerate training by normalizing activations and reducing internal covariate shifts.

2.2.3 Dense Layers: These layers aggregate extracted features and classify them into real or fake.

2.2.4 Dropout Layers: Randomly deactivates neurons during training, preventing overfitting by forcing the model to generalize better.

When using InceptionV3 as the feature extractor, the architecture consists of:

- Total Parameters: ~23.8 million
- Trainable Parameters: Varies depending on whether fine-tuning is enabled (e.g., when only training the classification head, it's significantly lower).
- Non-trainable Parameters: In a frozen feature extractor setting, the majority of InceptionV3's parameters (~23 million) remain fixed, contributing only to inference rather than learning new weights.

To enhance generalization and prevent overfitting, we incorporated transfer learning

2.3 Transfer Learning: By using pre-trained weights from MobileNet, the model learns robust features without requiring excessive training data. Transfer learning was implemented using pre-trained CNN architectures, specifically InceptionV3 (23.8M parameters, 21.8M trainable, 2M non-trainable) and InceptionResNetV2 (55.9M parameters, 54.3M trainable, 1.6M non-trainable). The convolutional base was initially frozen to retain learned features from ImageNet, preventing overfitting while reducing computational costs. Feature extraction was performed using the Global Average Pooling (GAP) layer, followed by a custom classifier comprising a 512-unit dense layer (ReLU), dropout (0.5), and a sigmoid output layer for binary classification. Fine-tuning involved unfreezing select higher-level convolutional layers to adapt domain-specific patterns. Overfitting mitigation strategies included data augmentation, dropout regularization, early stopping, and a learning rate scheduler to optimize convergence. This approach significantly improved classification accuracy and model generalization while maintaining efficient training on limited deepfake datasets.

By integrating these architectural choices and regularization techniques, our CNN-based deepfake detection model achieves high accuracy while maintaining computational efficiency, ensuring robustness across various datasets.

3. Experimental Results

3.1 Dataset

The dataset chosen to train the Deep Learning model is Deepfake Detection Challenge (DFDC) dataset. The dataset includes 590 actual YouTube videos with participants of various ages, ethnic groups, and genders, as well as Deepfake videos. The DFDC dataset used in this study is publicly available and can be accessed at https://www.kaggle.com/competitions/deepfake-detection-challenge/data.

The Deepfake Detection Challenge (DFDC) dataset was created to support the development of deepfake detection models by providing a large-scale collection of real and manipulated videos. It was introduced by Facebook, AWS, and other AI research organizations as part of a global initiative to combat the spread of AI-generated synthetic media. This dataset serves as one of the most comprehensive and diverse collections of deepfake videos available for training and testing machine learning models.

The DFDC dataset consists of over 100,000 videos, including both authentic and deepfake-generated clips. The deepfake videos were created using a variety of face-swapping and manipulation techniques, ensuring that models trained on this dataset can generalize well to different types of synthetic media. The dataset includes variations in lighting, facial expressions, angles, and backgrounds, making it highly diverse and representative of real-world conditions. We are considering the dataset being 400 real videos and 400 fake videos, allowing them to divide into training and testing data for model evaluation.

3.1.1 Files

- **train_sample_videos.zip** - a ZIP file containing a sample set of training videos and a metadata.json with labels. the full set of training videos is available through the links provided above.

- **sample_submission.csv** - a sample submission file in the correct format.
- **test_videos.zip** - a zip file containing a small set of videos to be used as a public validation set.

3.1.2 Columns

- **filename** - the filename of the video
- **label** - whether the video is REAL or FAKE
- **original** - in the case that a train set video is FAKE, the original video is listed here
- **split** - this is always equal to "train".

3.2 Pre-Trained Convolutional Neural Network Used for Feature Extraction

3.2.1 Pre-Trained Inception-v3 as a Feature Extractor

The input layer processes video frames resized to 299 × 299 × 3, while the output layer provides a classification decision on whether the frame is real or fake. From the input layer to the final 8×8×2048 max pooling layer, the model serves as a feature extractor, capturing essential spatial and texture information. The extracted 2048-dimensional feature vector represents high-level features such as edges, textures, and inconsistencies introduced by deepfake manipulations. In this work, InceptionV3 is utilized to extract discriminative features from video frames, which are then passed to a secondary classification network to detect subtle artifacts indicative of forgery.

3.2.2 Pre-Trained Inception-Resnet-V2 as a Feature Extractor

The Inception-ResNet-v2 model processes video frames resized to 299 × 299 × 3 as input, while the final classification layer determines whether a frame is real or fake. The network integrates Inception modules with residual connections, enabling efficient feature extraction while preserving crucial spatial and structural information. The extracted 1536-dimensional feature vector from the 8×8×1536 max pooling layer encapsulates fine-grained patterns and deep hierarchical features, making it well-suited for detecting manipulated content. In this work, Inception-ResNet-v2 is leveraged to extract high-level representations from video frames, which are then passed through a secondary classification model to differentiate authentic footage from deepfake-generated content based on subtle inconsistencies.

4. Comparative Analysis:

Table 2: Overall Detection Comparative Analysis

Pre-Trained Models	**Classifier**	**Accuracy (in %)**
InceptionV3	Convolutional Neural Network	81.2
InceptionResNetV2	Convolutional Neural Network	79.9

- Accuracy: The proportion of correctly identified real and fake images.

- The performance of pre-trained deep learning models in deepfake detection is evaluated using InceptionV3 and Inception-ResNet-v2, both serving as feature extractors followed by a Convolutional Neural Network (CNN) classifier. InceptionV3 achieves an accuracy of 81.2%, demonstrating its effectiveness in capturing fine-grained spatial and textural details, which are critical for identifying deepfake artifacts.
- On the other hand, Inception-ResNet-v2 attains an accuracy of 79.9%, slightly lower than InceptionV3. This may be attributed to its deeper architecture and reliance on residual connections, which, while enhancing gradient flow and feature reuse, might introduce subtle trade-offs in detecting fine-scale manipulations.
- The results suggest that InceptionV3's balance between depth and computational efficiency makes it more effective for deepfake detection, while Inception-ResNet-v2 remains a strong alternative for extracting high-level representations in complex scenarios.

Table 3 Classification using Convolutional Neural Network with InceptionV3

Detection	Precision (in %)	Recall (in %)	F1-Score (in %)	Accuracy (in %)
Real	82.84	84.00	83.42	81.3
Fake	81.30	87.00	84.10	83.5

The classification performance of InceptionV3 combined with a Convolutional Neural Network (CNN) for deepfake detection is analyzed through key evaluation metrics, including precision, recall, F1-score, and accuracy.

The model achieves 82.84% precision and 84.00% recall for real videos, leading to an F1-score of 83.42%, indicating a strong ability to correctly identify authentic content. For deepfake detection, the model attains 81.30% precision and 87.00% recall, resulting in an F1-score of 84.10%, showcasing its effectiveness in capturing subtle artifacts introduced by forgery techniques. The overall accuracy for real and fake classifications stands at 81.3% and 83.5%, respectively, reinforcing the model's robust feature extraction and classification capabilities.

These results highlight InceptionV3's ability to balance precision and recall, making it a reliable architecture for detecting manipulated media in deepfake detection tasks.

5. Results and Discussions:

The deepfake detection model, incorporating InceptionV3 and a custom CNN architecture, was trained and evaluated using the Deepfake Detection Challenge (DFDC) dataset. The results demonstrate a high level of accuracy in distinguishing between real and fake videos, showcasing the effectiveness of deep learning-based approaches in deepfake identification.

During model evaluation, the InceptionV3-based model on the test set, leverages its deep feature extraction capabilities. Meanwhile, the custom CNN model, designed with multiple convolutional and pooling layers, achieved an accuracy of around 81%, proving to be an efficient alternative with fewer computational requirements. The training and validation loss curves showed steady convergence, indicating that the models were well-optimized.

However, challenges remain, such as improving detection against highly sophisticated deepfakes and ensuring real-time processing efficiency. Future work will focus on fine-tuning model hyperparameters, incorporating adversarial training, and exploring multi-modal analysis to further enhance deepfake detection capabilities.

6. Conclusion:

In conclusion, deepfake detection plays a crucial role in safeguarding digital authenticity and preventing the spread of manipulated media. This project integrating InceptionV3 alongside a custom CNN model to develop a robust deepfake detection system. The system classifies real and fake videos by capturing intricate facial distortions and inconsistencies.

Our findings highlight the effectiveness of deep learning models in distinguishing deepfakes with high accuracy, demonstrating the potential of AI-driven detection mechanisms in mitigating digital misinformation. The combination of a pre-trained model and a custom CNN allows for both generalization across diverse deepfake manipulations and the flexibility to adapt to evolving threats.

Despite these advancements, deepfake technology continues to evolve, presenting challenges that require continuous improvements in detection methodologies. Future research can focus on enhancing model robustness against adversarial attacks, reducing computational complexity for real-time applications, and integrating multi-modal detection techniques that analyze both visual and audio inconsistencies. By strengthening deepfake detection methods, we can contribute to a safer digital ecosystem, ensuring the integrity of media content across various domains, including journalism, cybersecurity, and forensic investigations.

7.References:

[1] Aurchana, P., Arieth, R. M., & Revathy, G. (2024, July). Ensemble Based Soil Classification Using Machine Learning Techniques. In 2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing

[2] Aurchana, P., Revathy, G., Theodore, S. K., Renugadevi, A. S., Sesadri, U., & Vadivukarassi, M. (2023, December). Machine Leaming Technique for Deteching Leaf Disease. In International Conference on Advancements in Smart Computing and Information Security (pp. 30-39). Cham: Springer Nature Switzerland

[3] Revathy, G., Aurchana, P., Logeshwari, P., Priya, P. M., & Kalaiselvi, L. (2022, March). Mouth Gesture Classification using Computational Intelligence. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1424-1427). IEEE.

[4] Aurchana, P. K., & Prabavathy, S. (2021). Musical instruments sound classification using GMM. London Journal of Social Sciences, (1), 14-25.

[5] Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE access, 10, 25494-25513.

[6] Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). The deepfake detection challenge (dfdc) dataset. arXiv preprint arXiv:2006.07397.

[7] Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(2), e1520.

[8] Zhou, Y., & Lim, S. N. (2021). Joint audio-visual deepfake detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 14800-14809).

[9] Nie, F., Ni, J., Zhang, J., Zhang, B., & Zhang, W. (2024). DIP: Diffusion Learning of Inconsistency Pattern for General DeepFake Detection. IEEE Transactions on Multimedia.

[10] Masood, M., Nawaz, M., Javed, A., Nazir, T., Mehmood, A., & Mahum, R. (2021, May). Classification of Deepfake videos using pre-trained convolutional neural networks. In 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2) (pp. 1-6). IEEE.
[11] Sundaram, V., Senthil, B., & Vekkot, S. (2024, June). Enhancing Deepfake Detection: Leveraging Deep Models for Video Authentication. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.
[12] Guefrechi, S., Jabra, M. B., & Hamam, H. (2022, May). Deepfake video detection using InceptionResnetV2. In 2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-6). IEEE.
[13] Zhang, L., Zhao, D., Lim, C. P., Asadi, H., Huang, H., Yu, Y., & Gao, R. (2024). Video deepfake classification using particle swarm optimization-based evolving ensemble models. Knowledge-Based Systems, 289, 111461.
[14] Chhikara, R., & Punyani, P. (2024, May). Exploring Deepfake Detection: A Comparative Study of CNN Models. In 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS) (pp. 1-6). IEEE.
[15] Aurchana, P., & Dhanalakshmi, P. Deep Learning on Histopathological Images: Automated Classification of Oral Squamous Cell Carcinoma Stages Detection using Pre-trained Convolutional Neural Networks