DEEP FAKE VIDEO DETECTION USING INCEPTION RESNETV2

[1]Chinthaginjala Shalini, [2]Chilaka Thanuja
[1,2]UG Student, [1,2]Department of Computer Science & Engineering, Geethanjali Institute of Science and Technology, Gangavaram, Andhra Pradesh, India

ABSTRACT
Deepfake technology is widely used, which has led to serious worries about the authenticity of digital media, making the need for trustworthy deepfakeface recognition techniques more urgent than ever. This project employs a resource-effective and transparent cost-sensitive deep learning method to effectively detect deepfake faces in videos. To create a reliable deepfake detection system, A Convolutional Neural Network (CNN) model called InceptionResNetV2, were used. FaceForensics++ as benchmark datasetswere used to assess the performance of our method. To efficiently process video data, key frame extraction was used as a feature extraction technique. Our main contribution is to show the model's adaptability and effectiveness in correctly identifying deepfake faces in videos. A deep learning model is proposed to address this challenge, achieving over 90% accuracy in distinguishing real from fake videos. However, continuous improvement is necessary as both deepfake generation and detection techniques evolve.

INTRODUCTION
Deepfakes, AI-manipulated videos, are becoming a major threat. They can be used to spread misinformation and damage reputations. This project aims to combat deepfakes using InceptionResNetV2, a powerful AI model. This model can identify the subtle inconsistencies that give away a deepfake, such as unnatural facial features or vocal anomalies.
This project is important because deepfakes can erode trust in media, sow discord in elections, and damage reputations. By equipping platforms with deepfake detection tools, we can empower them to flag suspicious content. News outlets and individuals can also benefit from this technology.
In the following sections, we'll explore how InceptionResNetV2 works, how it'strained to identify deepfakes, and how its effectiveness is measured. This project aims to build a future where deepfakes are exposed, making the online world more secure and trustworthy.

MOTIVATION
The motivation behind this project stems from the urgent need to safeguard our digitallandscape from the threats posed by deepfakes. As deepfake technology becomes moresophisticated, the ability to distinguish real from artificial becomes increasingly difficult for the untrained eye. This can have significant consequences, eroding trust in media sources andcreating a climate of uncertainty.
By developing a reliable deepfake detection system, we aim to empower individuals and institutions to critically evaluate the information they encounter online. Imagine a world where social media platforms can automatically flag potential deepfakes, or news outlets can verify the authenticity of videos before publishing them. This project contributes to building such a future.
InceptionResNetV2 offers a promising avenue for deepfake detection due to its strengths in feature extraction and image classification. This pre-trained model has been shown to excel in various computer vision tasks, making it a suitable candidate for identifying the subtle inconsistencies often present in deepfakes. Our project will explorehow to fine-tune InceptionResNetV2 to recognize the specific patterns indicative of deepfake manipulation.
This work holds significant value for various stakeholders. Content creators canbenefit from having their work protected from being misused in deepfakes. Journalists and researchers can leverage this tool to enhance the credibility of their information sources. Ultimately, society as a whole stand to gain from a more

secure and trustworthy digital environment.

MACHINE LEARNING

Machine Learning a branch of artificial intelligence that develops algorithms by learning the hidden patterns of the datasets used it to make predictions on new similar type data, without being explicitly programmed for each task.

Traditional Machine Learning combines data with statistical tools to predict an output that can be used to make actionable insights.

Machine learning is used in many different applications, from image and speech recognition to natural language processing, recommendation systems, fraud detection, portfolio optimization, automated task, and so on. Machine learning iswidely applicable across many industries. Recommendation systems for example, are used by e-commerce, social media and news organizations to suggest content based on a customer's past behaviour.

Machine learning algorithms and machine vision are a critical component of self-driving cars, helping them navigate the roads safely. In healthcare, machine learning is used to diagnose and suggest treatment plans. The discipline of machine learning employs various approaches to help computers learn to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid.

This can then be used as training data for the computer to improve the algorithms ituses to determine correct answers.

OBJECTIVE

The objective of the "Deepfake Video Detection Using InceptionResNetV2" project is to develop a robust and accurate system for identifying deepfakes – synthetic videos manipulated using artificial intelligence. Deepfakes pose a significant threat to our digital landscape, eroding trust in media sources, fueling misinformation campaigns, and potentially damaging reputations.

This project aims to address this challenge by leveraging the power of InceptionResNetV2, a state-of-the-art deep learning architecture. We will train this model to recognize the subtle inconsistencies inherent in deepfakes, such as unnatural skin textures, misplaced shadows, or inconsistencies in lip movement during speech. By meticulously analysing video frames, InceptionResNetV2 will learn to differentiate between authentic videos and those manipulated with deepfake techniques.

The primary goal is to achieve high accuracy in detecting deepfakes. This involves training the model on extensive datasets containing both real and deepfake videos. The model will be fine-tuned to identify the minute details that expose deepfakes, ensuring it caneffectively distinguish between genuine and fabricated content.

Literature Review

Deepfake Detection Using XceptionNet:

"Ashok V, Preetha Theresa Joy"

This research proposes a method for detecting deepfakes, which are increasingly accessible and pose a threat to society. Existing methods struggle with high-quality deepfakes. The approach uses a state-of-the-art neural network called Xception to identify patterns and anomalies in images and videos that indicate manipulation. This network is trained on a large dataset of real and deepfake content to ensure it can accurately classify even unseen deepfakes. The goal is to create a reliable system to distinguish real content from deepfakes, protecting theintegrity of information and media.

EfficientNetV2: Smaller Models and Faster Training:

"Mingxing Tan, Quoc V. Le"

This research introduces EfficientNetV2, a new group of convolutional neural networks designed to train faster and with better efficiency than previous models. The development process involved a combination of searching for neural network architectures that are aware of training requirements, and also scaling them to find an optimal balance between training speed and parameter efficiency. These efficient networks were built

using a search space that included new operations like Fused-MBConv. Experiments showed that EfficientNetV2 models train significantly faster than current leading models, while requiring up to 6.8 times fewer parameters.

Deepfake Detection Using the Rate of Change between Frames Based onComputer Vision:

"Gihun Lee, Mihui Kim"

AI has become a powerful tool in various fields but also raises security, privacy, and ethical concerns. Deepfakes, realistic AI-generated fake videos, can be misused for malicious purposes. This paper proposes a new method to detect manipulated videos by analyzing changes in computer vision features between frames. This method achieved a 97% detectionrate, outperforming existing techniques, and remained effective even against methods designed to fool machine learning detection. This research offers a promising approach to combat deepfakes and safeguard the integrity of online video content.

Deepfake Detection Using SVM:

"Harsh Agarwal, Ankur Singh, Rajeswari D"

Due to the increasing realism of deepfakes created by generative networks, there's a growing concern about their potential misuse, such as spreading misinformation or blackmail. This hasled to a surge in efforts to detect deepfakes. This research proposes a method using Support Vector Machines (SVM) to identify deepfakes. The method involves analyzing videos in the frequency domain to find unnatural features invisible to the human eye. The researchers evaluate their approach on a dataset of deepfake videos and report promising results for detecting these manipulated videos.

DeepFake Face Image Detection based on Improved VGG ConvolutionalNeural Network:

"Xu Chang; Jian Wu, Tongfeng Yang, Guorui Feng"

This research proposes a novel deepfake detection method using an enhanced VGG network named NA-VGG. Deepfakes pose a significant challenge due to their realism. NA-VGG tackles this issue by analyzing image noise, a potential indicator of manipulation. The methodextracts subtle noise features and weakens facial features within the image to train the NA- VGG network to distinguish between real and deepfake images. Experiments showed that NA-VGG outperforms existing techniques, suggesting its potential as a powerful tool in the fight against deepfakes.

EXISTING SYSTEM

Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification aswell as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

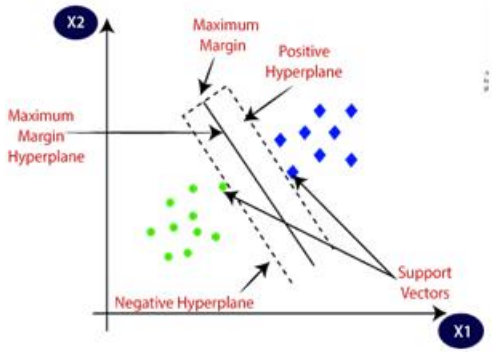

Fig.1. Support Vector Machine

Frame Exchange Rate Analysis

Frame Exchange Rate Analysis (FERA) is a technique used to detect deepfakes by analyzing variations in the consistency of video frames. Deepfake creation often involves manipulating individual frames or splicing content from different sources. FERA can potentially expose these inconsistencies by examining how the statistical properties of video frames change over time.

Here's a breakdown of how FERA works:

Frame Preprocessing: Each video frame is preprocessed to extract relevant features. This might involve converting the frame to grayscale, calculating histograms, or applying edge detection algorithms.

Feature Comparison: FERA compares features extracted fromconsecutive frames. Statistical measures like standard deviation or entropy can be used to quantify these variations.

Anomaly Detection: Large or unexpected changes in feature values between frames might indicate potential manipulation. Statistical thresholds or machine learning models can be used to identify these anomalies as potential signs of a deepfake.

PROPOSED SYSTEM

In this project we propose InceptionResNetV2, a powerful deep learning architecture,as the core of our deepfake video detection system.

InceptionResNetV2 excels at feature extraction and image classification, making it a suitable candidate for identifying the subtle inconsistencies that often betray a deepfake. Imagine the model scrutinizing a video, looking for unnatural smoothness in someone's face, inconsistencies in lighting, or minute glitches in lip movement during speech. These are the fingerprints of manipulation that InceptionResNetV2 will be meticulously trained to recognize.

The effectiveness of our system hinges on training InceptionResNetV2 on extensive datasets containing both authentic and deepfake videos. By meticulously analyzing vast amounts of video data, the model will learn the nuances that differentiate real from fabricated content. The training process involves fine-tuning the model's internal parameters to make it a deepfake detection expert.

This project goes beyond just achieving high accuracy. We aim to ensure the generalizability of the system – its ability to function effectively across diverse video scenarios. This involvestesting the model on a wide range of video samples, ensuring it maintains accuracy even when encountering new deepfake creation techniques.

By deploying InceptionResNetV2 as the foundation of our deepfake detection system, we aspire to contribute to a more secure and trustworthy digital space. This technology can empower platforms to flag suspicious content, allowing users to make informed judgments about the information they encounter online.

SYSTEM ARCHITECTURE

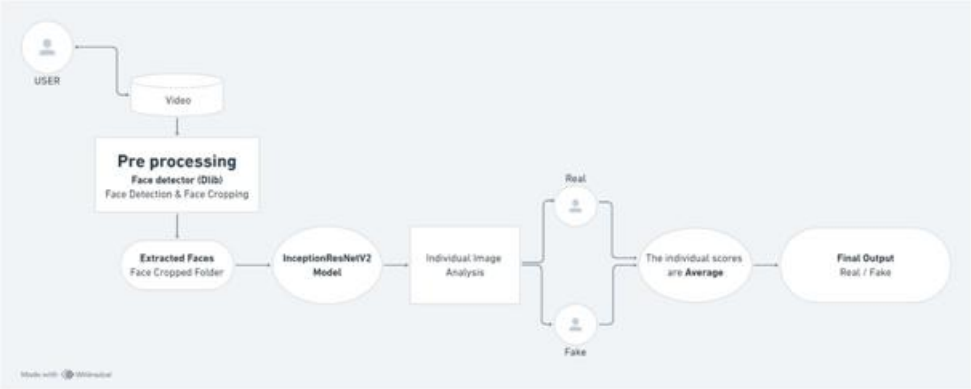

Fig.2. System Architecture

Preprocessing and Face Detection:

Input: You start with the video you want to analyze.

Frame Extraction: The system breaks down the video into individual frames (imagesrepresenting each moment in time).

Face Detection: In each frame, a face detector (like Dlib) identifies and isolates anyhuman faces present.

Face Cropping: The detected faces are cropped from their respective frames,resulting in separate images.
Saving Faces: All the cropped face images are saved in a designated folder.
Deepfake Detection with Inception ResNet V2:
Image Input: The folder containing the cropped faces becomes the input for thisstage.
Inception ResNet V2: Each image is fed into the Inception ResNet V2 model, a pre-trained deep learning architecture known for image classification.
Individual Image Analysis: The model analyzes each image independently,searching for patterns or inconsistencies indicative of deepfake manipulation.
Deepfake vs. Real Classification: Inception ResNet V2 outputs a probability scorefor each image, indicating the likelihood of it being a deepfake or a real face.
Video Classification (Final Verdict):

- **Accuracy Scores:** The system gathers the individual probability scores assigned by Inception ResNet V2 for each face image.
- **Averaging Accuracy:** To account for variations within the video, the individual scores are averaged.
- **Thresholding:** A predefined threshold value is set (e.g., 0.5).
- Final Classification:
  - **Real Video:** If the average score is below the threshold, the final output isclassified as a real video (low likelihood of deepfakes).
  - **Deepfake Video:** If the average score exceeds the threshold, the final output isclassified as a deepfake video (high likelihood of manipulation).

INCEPTIONRESNETV2
InceptionResNetV2: A Convolutional Neural Network for Image Classification
InceptionResNetV2 is a convolutional neural network (CNN) architecture developed by Google for image classification. It combines elements from two successful architectures: Inception and ResNet.
Inception modules: These modules process information at different scales simultaneously, capturing more diverse features from the input image.
Residual connections: These connections skip over some layers in the network, alleviatingthe vanishing gradient problem and allowing for deeper networks with better performance.
InceptionResNetV2 has achieved state-of-the-art results on various image classification benchmarks, demonstrating its effectiveness in extracting meaningful features from images. **Deepfake Video Detection with InceptionResNetV2**
Deepfakes are realistic manipulated videos created using machine learning techniques.
Detecting deepfakes is crucial for mitigating their harmful effects, such as the spread ofmisinformation and identity theft.
InceptionResNetV2 can be used for deepfake video detection in several ways:

- **Feature extraction**: The network can extract features from video frames, focusing on facial features, skin texture, and other visual cues that might reveal inconsistencies in manipulated videos.
- **Temporal analysis**: By analyzing features from multiple frames sequentially, InceptionResNetV2 can identify unnatural motion patterns or inconsistencies in blinking, often present in deepfakes.
- **Transfer learning**: A pre-trained InceptionResNetV2 model can be fine-tuned on a datasetof real and fake videos, leveraging its learned features for more efficient deepfake detection.

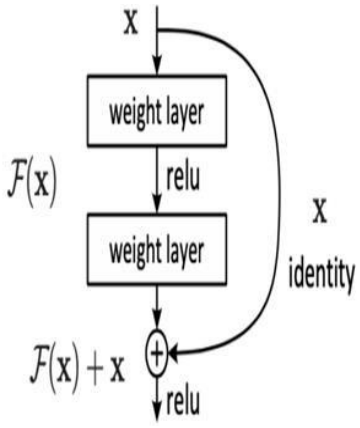

Fig.3. Residual connection

SYSTEM MODULES

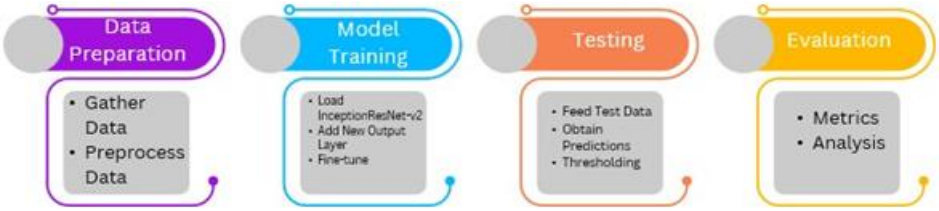

Fig.4. System Modules

Data Preparation:

Gather Data**:** Collect a balanced dataset of real and deepfake images/videos. Ideally, the dataset should represent diverse demographics, lighting conditions, and deepfake creation techniques.

Preprocess Data**:** Normalize and resize images/videos to the model's inputrequirements. Extract facial regions using techniques like dlib face detector if focusing on facial manipulation.

Model Training:

Load InceptionResNet-v2**:** This pre-trained model acts as a feature extractor, identifyingpatterns in images. Remove the original output layer.

Add New Output Layer**:** Append a single neuron with sigmoid activation, classifying imagesas real (0) or deepfake (1).

Fine-tune**:** Train the model using the prepared dataset. This adjusts the weights of the pre-trained layers to specialize in deepfake detection.

Testing:

Feed Test Data**:** Input new images/videos to the trained model.

Obtain Predictions**:** The model generates probabilities for real and deepfake classes.

Thresholding**:** Choose a threshold (e.g., 0.5) to categorize predictions as real (probability <threshold) or deepfake (probability >= threshold).

Evaluation:

Metrics**:** Calculate metrics like accuracy, precision, recall, and F1-score to assess the model'sperformance on unseen data.

Analysis**:** Analyse false positives and negatives to understand model limitations and potentialimprovements.

RESULTS

The execution of the process will be explained clearly with the help of continuousscreenshots.

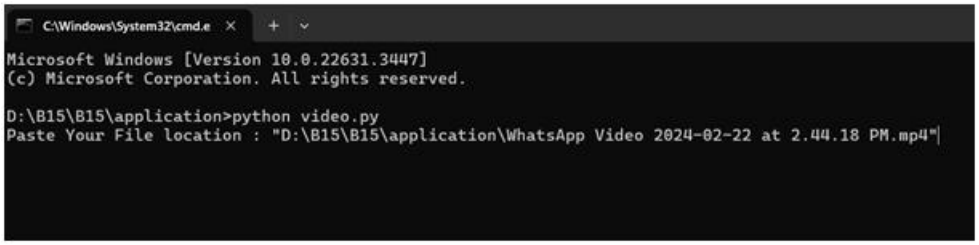

Fig.5.Video File Location

Fig.6. Crop the detected Faces

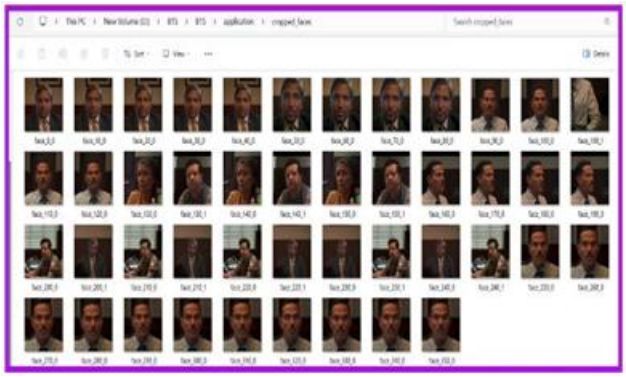

Fig.7. Store the Cropped faces in a Folder

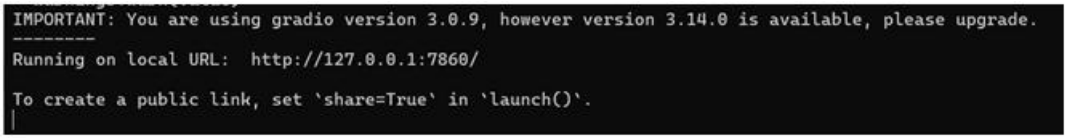

Fig.8. After execute the main file, it gives the Local URL (website)

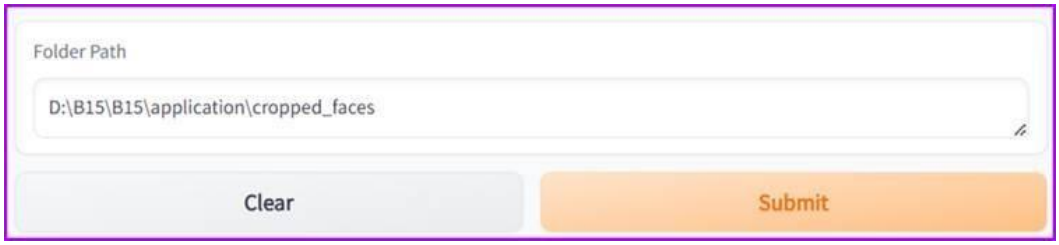

Fig 9. Folder Path (cropped faces)

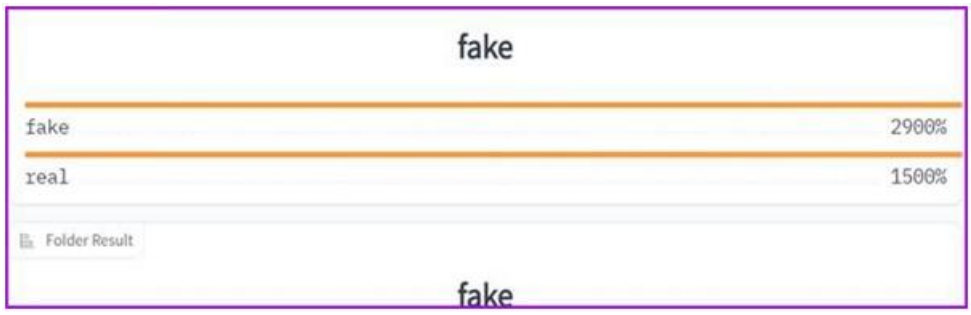

Fig.10. Individual Face Analysis

CONCLUSION

This project investigated the potential of InceptionResNetV2 for detecting deepfakes in videos using the FaceForensics++ dataset. We implemented a system that first preprocessed the video by extracting frames and identifying faces. Each cropped face image was then fed into the pre-trained InceptionResNet V2 model. The model assigned a probability score to each image, indicating the likelihood of it being a deepfake. Finally, by averaging theindividual scores and applying a threshold, we classified the entire video as real or deepfake. This approach leveraged the ability of InceptionResNet V2 to recognize subtle inconsistencies often present in deepfaked faces. The success of this project demonstrates the potential of deep learning models for combating the spread of deepfakes.

FUTURE SCOPE

This project lays the groundwork for further development. To enhance detection capabilities, we can explore training the model on a richer dataset encompassing a wider variety of deepfakes. Additionally, combining InceptionResNetV2 with other deep learning modelsthrough ensemble learning holds promise for potentially improving accuracy. Furthermore, incorporating explainable AI (XAI) techniques would provide valuable insights into the model's decision-making process, allowing for targeted improvements and increased user trust. Finally, optimizing the system for real-time processing would make it significantlymore practical for real-world applications. By focusing on these future considerations, we cancontribute to a more robust and reliable solution for identifying deep fakes.

References

1. W. M. Wubet, “The deepfake challenges and deepfake video detection,” Int. J. Innov. Technol. Explor. Eng, vol. 9, 2020.
2. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, “A machine learning based approach for deepfake detection in social media through key video frame extraction,”SN Computer Science, vol. 2, pp. 1–18, 2021.
3. M. Tanvir Rouf Shawon, G. Shahariar Shibli, F. Ahmed, and S. K Saha Joy, “Explainable cost-sensitive deep neural networks for brain tumor detection from brain mri images considering data imbalance,” arXiv e-prints, pp. arXiv–2308, 2023.
4. Z.-H. Zhou and X.-Y. Liu, “Training cost-sensitive neural networks with methods addressing the class imbalance problem,” IEEE Transactions on knowledge and dataengineering, vol. 18, no. 1, pp. 63–77, 2005.
5. G. Lee and M. Kim, “Deepfake detection using the rate of change between frames based on computer vision,” Sensors, vol. 21, no. 21, p. 7367, 2021.
6. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “Faceforensics++: Learning to detect manipulated facial images,” in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1–11.
7. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, “The deepfake detection challenge (dfdc) dataset,” arXiv preprint arXiv:2006.07397, 2020

8. Xu, J. Liu, J. Liang, W. Lu, and Y. Zhang, “Deepfake videos detection based on texture features.” Computers, Materials & Continua, vol. 68, no. 1, 2021.
9. P. Korshunov and S. Marcel, “Deepfakes: a new threat to face recognition? assessment and detection,” arXiv preprint arXiv:1812.08685, 2018.
10. Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-df (v2): a new dataset for deepfake forensics [j],” arXiv preprint arXiv, 2019.
11. Kohli and A. Gupta, “Detecting deepfake, faceswap and face2face facial forgeries using frequency cnn,” Multimedia Tools and Applications,vol. 80, pp. 18461–18478,2021.
12. Kim, S. Tariq, and S. S. Woo, “Fretal: Generalizing deepfake detection using knowledge distillation and representation learning,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 1001–1012

13. F. Chollet, “Xception: Deep learning with depthwise separable convolu tions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), July 2017.