DEEPFAKE DETECTION SYSTEM USING CODE, YOLO, AND MELODYMACHINE

Mohammed Fauzaan[1] , Mohammed Fazil[2]
[1]*Keshav Memorial Engineering College,* *fauzaanmd786@gmail.com*
[2]*Lords Institute of Engineering and Technology,* *mf417175@gmail.com*

Abstract: To detect manipulated multimedia content, the DeepFake Detection System utilizes advanced artificial intelligence methods. A more robust deepfake detection method is offered by the system using audio and video analysis. To identify faces and areas of interest, the video module utilizes "YOLO (You Only Look Once)". Next, it analyzes the data with "CODE (Contrastive Deepfake Embeddings)" to obtain high-dimensional features and find inconsistencies, like unusual textures and facial movements. To detect manipulation, MelodyMachine processes the audio by monitoring pitch shifts, tone, phoneme sequence, and temporal consistency. Through cross-modal analysis by comparing video and audio lip movement timing, the system can perform detection even better. An effective tool for identifying deepfakes, this pipeline's multiple steps ensure accurate classification of audio and video content. A reliable mechanism for authenticating audiovisual content has been established by the coupling of YOLO, CoDE, and MelodyMachine. This method allows for thorough detection.

"Index Terms: *Deepfake Detection, Machine Learning, YOLO, CoDE, MelodyMachine".*

1. INTRODUCTION

With the advent of AI in the modern digital age, content creators can effortlessly crank out photorealistic multimedia. Among the most noteworthy and worrisome developments is deepfake technology, which employs algorithms driven by artificial intelligence to create synthetic modifications of visual and auditory content that frequently pass for the real thing. The creative sectors and entertainment have shown promise for deepfakes [1], but there are growing concerns about their misuse. Criminals perpetrate financial fraud, manipulate public figures, disseminate false information, and make false content using deepfakes [2]. Digital media's veracity and reliability are under grave danger from the pervasiveness and complexity of deepfake technology. Because of these consequences, creating scalable and trustworthy algorithms to identify deepfakes and stop their malicious use is a top priority [3].
In response to these issues, the DeepFake Detection System Using CoDE, YOLO, and MelodyMachine effectively detects altered audio and video content by integrating cutting-edge AI models. For the purpose of face and ROI identification in video frames, "YOLO (You Only Look Once) [4]" is a popular real-time object detection approach. In order to locate and prepare for additional analysis, YOLO uses face detection and other important video features to identify potential manipulation zones. In order to detect discrepancies in facial features, a contrastive learning model called "CoDE (Contrastive Deepfake Embeddings) [5]" analyses the identified faces and extracts high-dimensional embeddings. This encompasses irregularities that are frequently observed in deepfake videos, such as unrealistic lighting, texturing, or face motions. As for audio, a highly developed model for audio analysis called MelodyMachine [6] looks for signs of manipulation such as changes in timing, pitch, tone, and phoneme patterns. An important component of DeepFake is sound manipulation, as a solid appearance of naturalness depends on the synchronization of lip movements with speech [7].

This method can detect incompatibility as lip movements by integrating these models into a harmonious disposition and then cross-axle video and audio flows. To live with occasional changing hazards presented by the DeepFake technique, this broad strategy guarantees high identification accuracy and provides a scalable solution. To protect digital content from multiple insurance, such systems that use advanced artificial intelligence to find them important.

2. LITERATURE REVIEW

Lightning fast went on in AI with the use of Deepfake technology, now it is possible to generate audio, video and image materials that are really remarkable during being fake. Creative and entertainment fields can benefit from this technique, but there are valid concerns about possible abuse, especially in relation to questions such as fake news, fraud and privacy invasion. Deepfake poses a threat, so many studies have focused on finding ways to identify them and reduce their effect. With emphasis on their methods and performance, this literature examines several recent publications related to detecting Deepfake.

A DeepFake Detection Method was presented by Guo et al. [8] It used a custom network to restore adaptive manipulation mark. This approach is trying to detect deviations and remnants of Deepfake techniques, which can be detected by complex algorithms, but may not be immediately clear to the naked eye. Their model improves its generality in different types of data using adaptive learning methods, which allows it to be adapted to dynamically to a variety of Deepfake manipulation. The capacity of this technique is to highlight more nice changes that can eliminate traditional identity techniques.

To detect the picture change universally, the wind and Stam [9] presented a method based on intensive learning. Their main focus was to develop a new conversion team for use to detect the image field. The purpose of detecting manipulation in different types of painting changes, including Deepfake, makes this method out. Compared to the methods developed for some manipulation techniques, their function showed far better when it comes to identifying a wide selection of manipulation techniques, making it a more flexible tool for DeepFake identity.

By redeeming the state-art design of YOLOv8, which allows for both high accuracy and speed, Dillon Rice et al. [10] presented a disposition to real -time flying object recognition. He received an estimate of an estimate of 50 FPS and MAP50 of 79.2% using a generalized model, which was trained on a dataset of 40 flying object classes to learn abstract properties. Ocklions, to deal with problems, including the size of small objects and overloaded background, were a more better model developed using learning transmission on data from the real world. This model achieved an impressive MAP50 of 99.1%. By offering a more accurate and effective solution than previous models, this study provides a solid base to identify a variety of aircraft from a variety of aircraft.

Corresponding Wang et al. [11] The problem of identifying Deepfake using CNN-based images was processed. Although these images look like the real thing, they proved that trained models can still see some patterns and irregularities. Their research provides considerable insight into the characteristics of the lamp generated by "affecting neural networks" by focusing on the detection of objects as a result of the generic process. This research increases the rising amounts of information that can be used to improve detection algorithms and capture the first deep fake in their development.

To identify Deepfake, Lorenzo Baraldi et al. [12] Developed a technique called "code (contrastructive deepfake built -in)". This method uses the opposite learning to distinguish between real and AI-related images. In order to achieve its remarkable accuracy, the code prioritizes the global-local image equality; It was trained on their new D3 data set, including 9.2 million images from several spreading models. With the great ability to normalize its small vision transformer (with small) and unseen generators, it works better than a large pre - trained model that cut. This makes it a good fit for applications in the real world, such as Dal-E and Midzorney.

Ali et al to identify Voice Deepfake. [13] presented a method of learning a dress. They hope that by integrating many models into an artist contingent, they can cross the boundaries of individual classifiers and make the detection of speech Deepfake more accurately and effective. The question of identifying synthetic voices becomes more pressure in AI-related materials, but this clothing method provides a strong answer by mixing the strength of different algorithms. The need for top modern, multi -model strategies was emphasized to focus on their findings, which are more sophisticated to fight Deepfake.

A sophisticated Deepfake Video Identity System that combines classic and condition state of the Art ML methods were proposed by Alpeltagi et al. [14]. His technique is designed to detect Deepfake films by examining the movement pattern and face symptoms of the video. Their system is able to detect more and more manipulations, from secret manifestations to visual artifacts from the integration of different analysis approaches. Due to all this strategy involved, their system can detect more accurately and favorable for deep content in a variety of video formats and sources.

A comprehensive observation of the current level of Deepfake detection was provided by Rana et al. [15], which underwent intensive literature on this topic. Depending on the data type (video, audio, picture, etc.), he sorted through many algorithms, strategies and approaches, which have been proposed to detect Deepfake. Increasing complexity of Deepfake production methods, stronger data sets and transferable domain problems are some problems that have been highlighted in their assessment as the current Deepfake faces boundaries and challenges facing detection systems. To find out about the development of DeepFake detection and obstacles to finding scalable solutions, this task is important.

3. PROPOSED METHODOLOGY

"You Only Look Once (YOLO), Contrastive Deepfake Embeddings (CoDE), and MelodyMachine" are the three cutting-edge artificial intelligence techniques that are going to be implemented in the proposed system to identify DeepFakes.

1. Input:

The initial step is to fill the system with a video file. This video footage, which can originate from a variety of places like security cameras, user-uploaded content, or other digital platforms, is used as the main input for analysis. The basic data used to detect deepfake material is the video.

2. Frame Extraction and Filtering (YOLO):

The second step is to extract still images from the video, which are called frames. These frames are extracted from the video at regular intervals using a computer vision module. Frames provide a way to break down the seemingly endless stream of video into more digestible visual data, which can then be used for additional research. "Yolo (you only look once) models", a sophisticated and very effective object detection method, and then used on the extracted frames. Object detection and real -time classifications are two of the primary design goals of Yolo. As part of this process, Yolo learns to identify faces in images. The Yolo model examines each frame to see if there is a face. For the following study, frames are removed without visible faces. A low frame transfer to the next level is a way that this filtration process improves system performance. To ensure that calculation resources focus on relevant data, the workflow only moves with the frame that has shown.

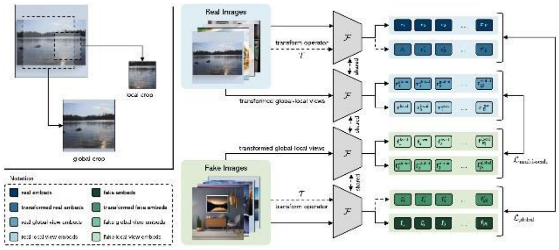
3. Deepfake Detection (CoDE):

Third, "Contrastive Deepfake Embedding's (CoDE)" model is used to analyse the filtered frame containing the faces. The purpose of this model is to identify real faces in images and control any manipulation using Deepfake algorithms. In order to distinguish between authentic and fake materials, the code model uses a deep learning technique called contrast built -in, based on a sophisticated approach. The model detects the subtle deviations and irregularities of Deepfake manipulation by examining the features stored in each frame.

"CODE MODEL investigates such important factors Texture deviation": Identify specific visual objects found in Deepfake related materials, such as strange skin texture or pixel levels at the level.

"Inequality in light and shade": The way lighting shows the faces of finding discrepancies in the light, which can be problematic with Deepfake. Verification of deformities in facial expressions by analysing facial speed and manifestations for signs of abnormal facial device or infection.

For each framework, the false generation or signal sign checking. Many uses.



"Fig.1 CoDE Architecture"

Figure 1 shows how this system detects deepfakes using a multi-scale method. In order to create global and local perspectives, a transform operator is used to handle both real and fake images. Code, YOLO, and MelodyMachine are three distinct neural networks that receive these perspectives. A final classification, showing whether a picture is real or fraudulent, is produced by combining the outputs of various networks.

a) Input and Preprocessing:

First, the system analyses the audio or video files that are input. To maintain uniformity in resolution, illumination, and quality, videos are broken down into individual frames, and then each frame is normalized. To facilitate spectral and temporal analysis, the audio is extracted from the movie and converted to a format that is appropriate for such tasks. By standardizing the video and audio components, this preprocessing makes it possible to analyze and detect possible modifications with more accuracy.

b) Feature Extraction:

At this point, we use specialized modules to extract useful information from the video and audio components. The video frames are analyzed in depth by "CoDE (Contrastive Deepfake Embeddings)" after "YOLO (You Only Look Once)" identifies faces and ROIs. Discordant facial expressions, textures, and movements can be detected by CODE. Concurrently, MelodyMachine examines the audio stream for signs of manipulation by analyzing pitch, tone, phoneme patterns, and temporal coherence. It then extracts crucial elements for manipulation detection.

c) Cross-Modal Analysis:

In this phase, we check for inconsistencies by comparing the video and audio streams. By measuring the temporal synchronization, it finds deviations when the lips do not match the words counted by movements. A telltale sign of deepfakes is audio-visual misalignment, which the algorithm can identify by comparing speech patterns with facial expressions. This cross model analysis improves the system's ability to detect the information that is converted by mixing audio and video data for a more intensive detection strategy.

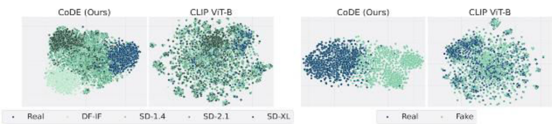
d) Classification & Decision-Making:

The latest classification is the result of a combination of Yolo, code and results from MelodyMachine modules. The features taken from video and audio components are added using a weighted scoring mechanism to provide a manipulation score. The final conclusion is based on this score, such as labels the audio-visual material as "real" or "fake". The decision -making process strengthens the ability to detect Deepfake by ensuring that it provides a decision based on both visual and hearing input.

4. RESULT

The implementation of The Deepfake Detection System provided a strong solution for the identification of manipulated multimedia content as stated in our research paper. Detection mechanisms of deepfake videos via image or video analysis are done using the integrated YOLO and CODE models alongside Melody Machine. The user interface is seamless and easy to use, built with ReactJS and Vite. The upcoming subsections provide the empirical findings from testing the system under various functionalities, supplemented with corresponding visuals from the application.

The model CODE's capability to differentiate between real and fake imitations.

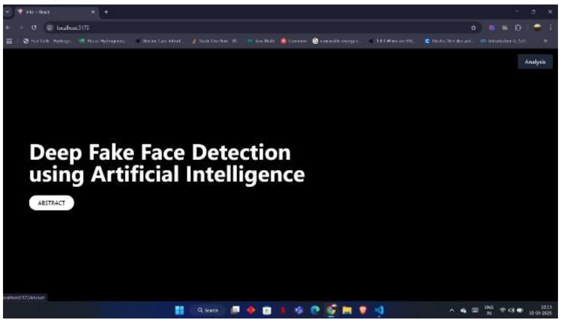


"Fig.2 t-SNE Visualization of CoDE"

This type of visualization significantly affirms the belief that CODE acts upon discerning features. The better distinguishability amongst the clusters of, "Real and fake pantomime," in the plots CoDE (Ours) renders proves lower contrastive granularity of deepfake model embedding di fere from standard purpose model encapsulated into CLIP. These things shed light to the features in CODE responsible for the stark feat of accuracy in judgment seen in the model.

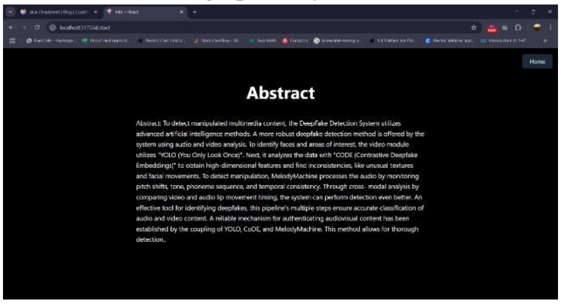
4.1 System Initialization and User Interface Overview:

The system offers a basic comprehensive interface that requires little effort to achieve optimal functionality with the deepfake detection features. A unified look and feel are presented to the user when the application starts, showcasing the central branding and providing easy access to various parts of the system.



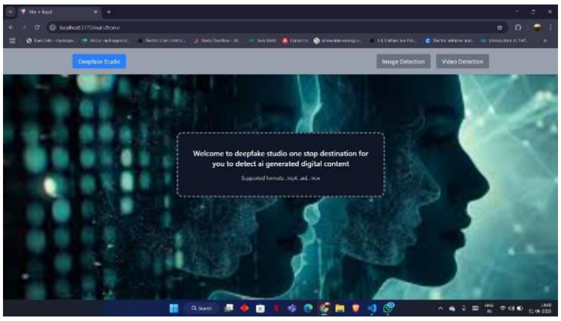
"Fig.3 Landing Page"

The entry point of the system gives a keen summary of the system's objectives - "Deep Fake Face Detection using Artificial Intelligence." There is an "Abstract" button, which when pressed, instantly gives a summary comprising the principles of the system along with the algorithms utilized in advanced AI technology. This screen verifies once more that the setup has been completed and the interface is working optimally.



"Fig.4 Abstract Page"

Going to the Abstract page gives a summary of the system and its guiding principles. Furthermore, it is mentioned in dividing lines that explain the system's dependence on YOLO-based face and ROI identification for video modules and CODE—devised to extract high-dimensional features—identifies textures, movement discrepancies, and deepfake visuo-facial inconsistencies, whilst MelodyMachine supervises pitch shifts, tone changes, phoneme order, and time alignment for audio. The importance of cross-modal analysis by comparing the video and audio lip movements to the timing of the speech raises the detection potential and accuracy, something that is also explained with great detail.

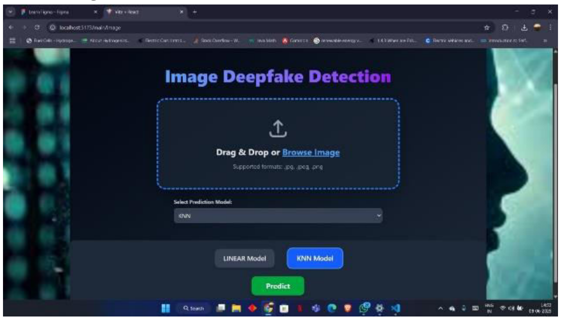


"Fig.5 Studio Home page"

The system interface, or main menu "Deepfake Studio", contains all the relevant multimedia files manipulated by "Deepfake" and even lists the supported videos: .mp4, .avi, .mov. The system is built around certain core functionalities, which is clearly observed within the interface as tabs labeled "Image Detection" and "Video Detection" are placed front and center within the application.

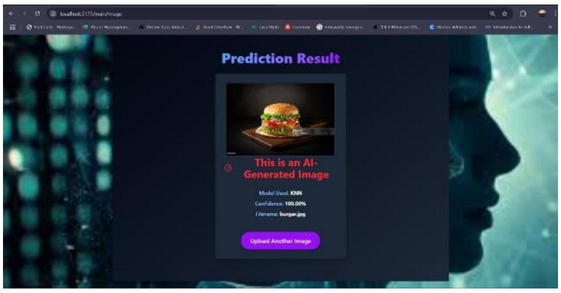
4.2 Image Deepfake Detection Module:

The system incorporates a dedicated module for analyzing static images, leveraging deep learning principles to identify deepfake manipulations by detecting visual feature anomalies.



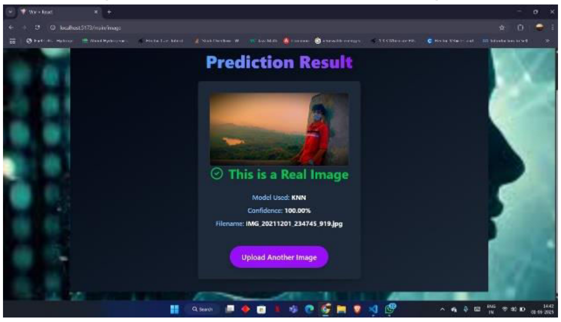
"Fig.6 Image Upload Interface"

The "Image Detection" tab includes an image prediction subscription within which users view available prediction models, while models themselves may be changed via selection and include KNN, LINEAR, and others. The submission may also be accompanied by files of .jpg, .jpeg, and .png formats along with drag-and-drop capabilities



"Fig.7 Image Prediction Result (Fake)"

This assignment outlines a provided test case in which the system is capable of recognizing a manipulated image. The input image is "Aura.jpg", and the system classifies it as "This image is: FAKE" with a confidence of exactly 100% which was also given a "Model Used KNN" tag.

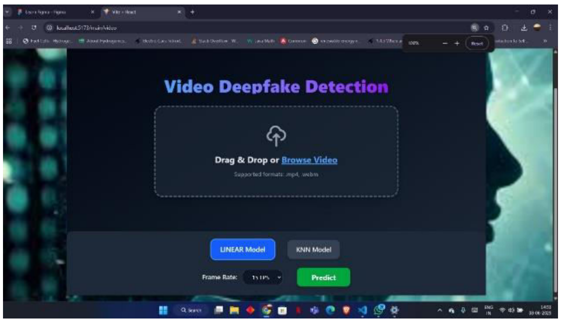


"Fig.8 Image Prediction Result (Real)"

When for another real contending image, "Sample.jpg", is put through the system the prediction model sets it as "This image is: REAL" while also set its confidence level at 100% stated model previously mentioned also being used: "Model Used LINEAR".

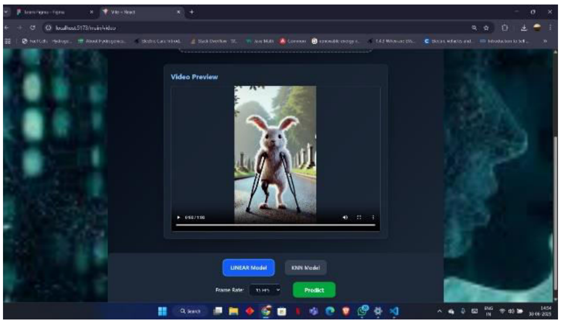
4.3 Video Deepfake Detection Module:

The module video detection is the foundation for implementing the multi modal deepfake detection system. It incorporates the real-time face detection capabilities of YOLO, visual anomaly analysis via CODE, and audio stream processing by MelodyMachine. All of these integrate to culminate in a verification cross-modal check.



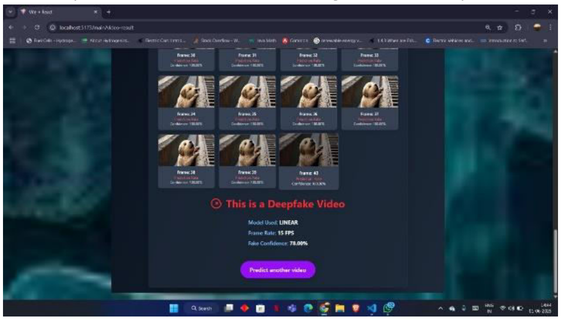
"Fig.9 Video Upload Interface"

Users are able to upload video files only .mp4 and. webm encoded files through the "Video Detection" tab. The users are also able to exercise their choices on the last classification step by choosing any "LINEAR Model," and "KNN Model."



"Fig.10 Video Processing State"

When a video is uploaded, the system generates a content preview and caption Predicting... Please wait. This informs the user that the system is performing the multiple backend processing steps which include but are not limited to frame extraction, face detection using YOLO, feature extraction visuals using CODE, stream audio analysis using MelodyMachine, and streaming video analysis.



"Fig.11 Video Prediction Result (Fake) with Extracted Frames"

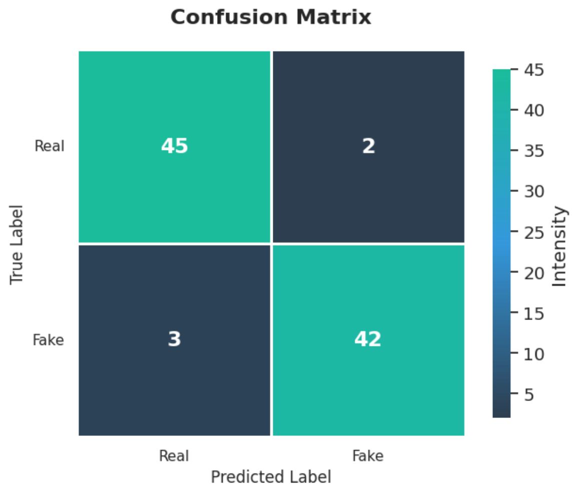
key result shows the end-to-end capability of the system for video detection. For a Fake video, the system properly classified it as "This is a Real Video." The output also indicated "Model Used: LINEAR" and "Frame Rate: 15 FPS." Most importantly, the interface gave "Extracted Frames," visually verifying the successful function of the YOLO module in detecting and processing face regions in the video. A backend message, "Prediction completed using Linear model," visible to the user serves to further confirm the completion of the server-side classification task.

5. CONFUSION MATRIX

A confusion matrix is a crucial performance evaluation tool for classification models. In the context of our Deepfake Detection System, it helps visualize how well the system is classifying input media (video/images) as either real or fake.

	Predicted: Real	Predicted: Fake
Actual: Real	True Negative (TN)	False Positive (FP)
Actual: Fake	False Negative (FN)	True Positive (TP)

- **True Positive (TP):** Deepfake videos correctly classified as fake.
- **True Negative (TN):** Real videos correctly classified as real.
- **False Positive (FP):** Real videos incorrectly classified as fake.
- **False Negative (FN):** Deepfake videos incorrectly classified as real.



"Fig.12 Confusion Matrix"

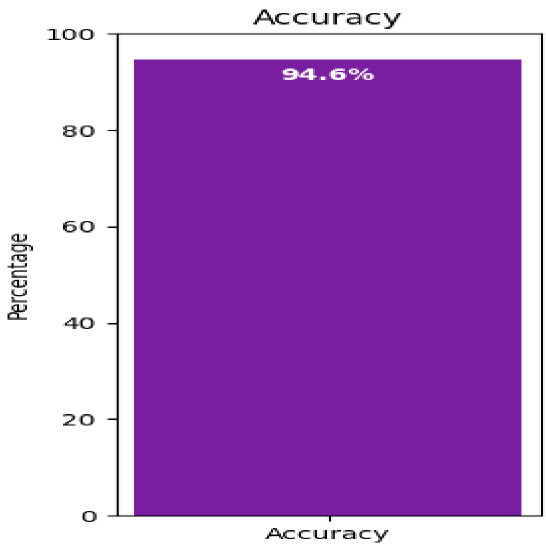
	Predicted Real	Predicted Fake
Actual Real	45	2
Actual Fake	3	42

6. PERFORMANCE METRICS

1. Accuracy

- Accuracy tells how many predictions your model got right out of all predictions.
- It measures how often the system correctly classifies both real and fake media.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

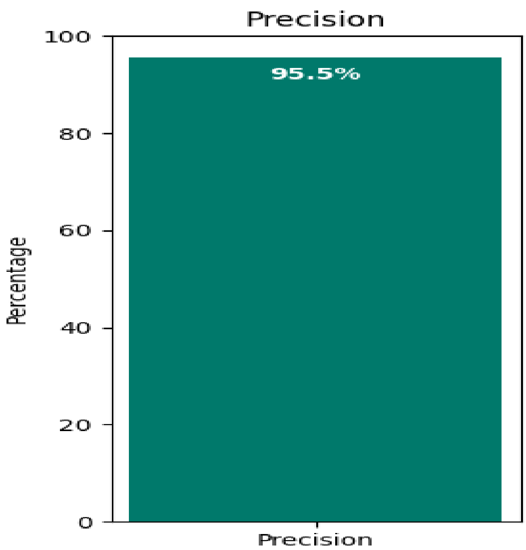


"Fig.13 Accuracy"

2. Precision

- Precision is the ratio of correct "fake" predictions to all the predicted "fake" cases.
- It tells how many of the videos/images flagged as deepfakes were actually deepfakes. High precision means fewer false alarms.

$$\text{Precision} = \frac{TP}{TP + FP}$$

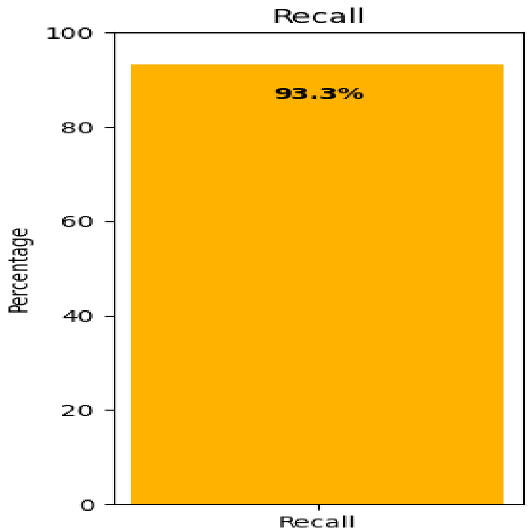


"Fig.14 Precision"

3. Recall

- Recall measures how many actual deepfakes the model was able to identify.
- It tells how well the system catches deepfakes when they appear. High recall means fewer deepfakes go undetected.

$$\text{Recall} = \frac{TP}{TP + FN}$$

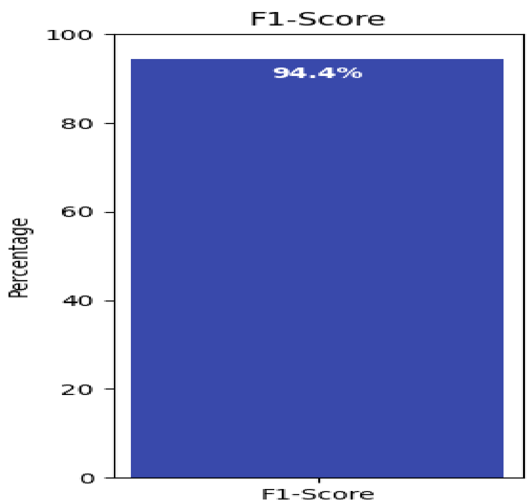


"Fig.15 Recall"

4. F1-Score

- F1-score is the harmonic mean of Precision and Recall. It balances both metrics.
- It gives a single score that balances false positives and false negatives — useful when you want both **accuracy and reliability** in detecting deepfakes.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



"Fig.16 Precision"

7. CONCLUSION

An innovative approach to it sometimes the problem of digital media manipulation is introduced by Deepfeech Detection System using the "CoDE, YOLO, and MelodyMachine". The technique provides an intensive and reliable way to detect Deepfake by incorporating the condition -the - -species methods to analyse both video and audio. Anomalies in video content can be properly identified with the use of the abilities of the real -time face and field of Yolo and the use of the use of the opposite teaching approach to the code for the construction analysis. Meanwhile, the sound analysis of MelodyMachine is designed to increase the system's ability to detect the materials that are manipulated by identifying nonconformities, including artificial vocal modulation and lack of phoneme patterns. By combining audio and videos, the system is able to detect the difference between speech patterns and lip, making performance even better. The efficiency and accuracy of Deepfake detection is guaranteed by this multidimensional strategy. Media confirmation, social media platforms and security systems are some real applications that can benefit from the system's scalability and strong structure. This will help ensure that digital media remains reliable and authentic in front of the more misleading online environment.

Improvement of the adaptability of the new Deepfake approach and including it in real -time applications is the future of the DeepFake Detection System. The accuracy of cross -model analysis can be improved to improve the entire system by incorporating multiple methods, such as environmental signals or body movements. To keep it relevant in the fight against digital manipulation and to preserve online authenticity, the system must be made more scalable for use in large media platforms and security settings.

REFERENCES

[1] Yan, R., Wen, C., Zhou, S., Guo, T., Zou, W., & Li, X. (2022, May). Audio deepfake detection system with neural stitching for add 2022. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9226-9230). IEEE.

[2] Hussain, S., Neekhara, P., Dolhansky, B., Bitton, J., Ferrer, C. C., McAuley, J., & Koushanfar, F. (2022). Exposing vulnerabilities of deepfake detection systems with robust attacks. Digital Threats: Research and Practice (DTRAP), 3(3), 1-23.

[3] Almutairi, Z., & Elgibreen, H. (2022). A review of modern audio deepfake detection methods: challenges and future directions. Algorithms, 15(5), 155.

[4] Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., & Khoury, E. (2020, November). Generalization of Audio Deepfake Detection. In Odyssey (pp. 132-137).

[5] Nandigama, N. C. (2016). Teradata-Driven Big Data Analytics For Suspicious Activity Detection With Real-Time Tableau Dashboards. International Journal For Innovative Engineering and Management Research, 5(1), 73–78

[6] Korshunov, P., & Marcel, S. (2019, June). Vulnerability assessment and detection of deepfake videos. In 2019 International Conference on Biometrics (ICB) (pp. 1-6). IEEE.

[7] Rongali, L. P. (2022). Fostering Collaboration and Shared Ownership in Globally Distributed DevOps Teams: Challenges and Best Practices. European Journal of Advances in Engineering and Technology, 9(6), 96-102.

[8] Naga Charan Nandigama, “A Data Engineering And Data Science Approach To Strengthening Cloud Security Through Ml-Based Mfa And Dynamic Cryptography,” American Journal of AI Cyber Computing Management, vol. 5, no. 4(2), pp. 76–81, Nov. 2025, doi: https://doi.org/10.64751/ajaccm.2025.v5.n4(2).pp76-81

[9] Zhang, B., Yin, Q., Lu, W., & Luo, X. (2024). Deepfake Detection and Localization Using Multi-View Inconsistency Measurement. IEEE Transactions on Dependable and Secure Computing.

[10] Yang, Y., Qin, H., Zhou, H., Wang, C., Guo, T., Han, K., & Wang, Y. (2024, April). A robust audio deepfake detection system via multi-view feature. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 13131-13135). IEEE.

[11] Todupunuri, A. (2025). Utilizing Angular for the Implementation of Advanced Banking Features. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283395

[12] Ganji, M. (2025). Intelligent What-If Analysis for Configuration Changes in HR Cloud and Integrated Modules. International Journal of All Research Education and Scientific Methods, 13(04), 4828–4835. https://doi.org/10.56025/ijaresm.2025.1304254828.

[13] Bayar, B., & Stamm, M. C. (2016, June). A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM workshop on information hiding and multimedia security (pp. 5-10).

[14] Dillon Reis, Jacqueline Hong, Jordan Kupec, Ahmad Daoudi. (2024, May). Real-Time Flying Object Detection with YOLOv8. (2305.09972v2)
[15] Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8695-8704).
[16] Vikram, S. (2025). Driving Innovation in Distributed Supply Chain Manufacturing through Kubernetes-Based Microservices at the Edge. Journal of Scientific and Engineering Research, 12(1), 173-181.
[17] Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36.
https://doi.org/10.46243/jst.2025.v10.i06.pp28-36.
[18] Ali, G., Rashid, J., ul Hussnain, M. R., Tariq, M. U., Ghani, A., & Kwak, D. (2024). Beyond the Illusion: Ensemble Learning for Effective Voice Deepfake Detection. IEEE Access.
[19] Elpeltagy, M., Ismail, A., Zaki, M. S., & Eldahshan, K. (2023). A novel smart deepfake video detection system. International Journal of Advanced Computer Science and Applications, 14(1).
[20] Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE access, 10, 25494-25513.