D.Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY, KOLHAPUR

(An Autonomous Institute)

DEPARTMENT OF CSE (DATA SCIENCE)

A

Project Report

on

"DeepScan Pro: A DeepFake Detector"

Submitted by

Name	Roll no.
Miss. Shreya Ramchandra Jadhav	10
Mr. Dhananjay Ramrao Ambatwar	12
Mr. Manasvi Harihar Mude	14
Mr. Aryan Shahaji Tapase	16

Under the guidance of

Mr. S. T. Powar

Final Year B. Tech. CSE (Data Science)

Academic Year 2024-25

D. Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY, KOLHAPUR

(An Autonomous Institute)

DEPARTMENT OF CSE (DATA SCIENCE)

CERTIFICATE

This is to certify that,

Roll No.	Unique ID	Student Name	Exam Seat No.
10	EN21123387	Miss. Shreya Ramchandra Jadhav	10149
12	EN21118971	Mr. Dhananjay Ramrao Ambatwar	10076
14	EN21175729	Mr. Manasvi Harihar Mude	10094
16	EN21172480	Mr. Aryan Shahaji Tapase	10066

have successfully completed the project work entitled,

“DeepScan Pro: A DeepFake Detector”

In partial fulfilment for the curriculum of **Final Year B. Tech. CSE (Data Science)**. This is the record of their work carried out during academic year 2024-2025.

Date:

Place: Kolhapur

Mr. S. T. Powar
Project Guide

Prof. DR. G. V. Patil
HOD

Prof. DR. S. D. Chede
Principal

External Examiner

DECLARATION

We the undersigned students of **Final Year B. Tech. CSE (Data Science)** declare that the project work report entitled "**DeepScan Pro: A DeepFake Detector"** written and submitted by us, under the guidance of **Mr. S. T. Powar,** is our original work. The empirical findings in this report are based on the data collected by us. The matter assimilated in this report is not the reproduction of any readymade report. We have not violated any of the provisions under the Copyright and Piracy / Cyber / IPR Act amended from time to time.

Date:

Place: Kolhapur

Roll No.	Unique ID	Student Name	Signature
10	EN21123387	Miss. Shreya Ramchandra Jadhav	
12	EN21118971	Mr. Dhananjay Ramrao Ambatwar	
14	EN21175729	Mr. Manasvi Harihar Mude	
16	EN21172480	Mr. Aryan Shahaji Tapase	

ACKNOWLEDGMENT

We extend our heartfelt gratitude to Prof. DR. A. K. Gupta Sir, Executive Director DYPCET, whose visionary leadership and unwavering support provided the foundation for this endeavour. We also express our sincere gratitude to Prof. DR. S. D. Chede Sir, Principal DYPCET, for their guidance and encouragement throughout this journey. Special thanks to DR. G. V. Patil Sir, Head of Department Data Science, for their invaluable insights and mentorship, shaping our understanding and approach within the field of data science.

We are immensely grateful to our project guide **Mr. S. T. Powar** whose expertise and dedication empowered us to navigate challenges and achieve milestones with confidence.

To our esteemed colleagues in the project group, your collaboration and dedication have been instrumental in realizing the goals of this project. Together, we have embraced innovation and teamwork, driving the project towards success. This project report stands as a testament to the collective efforts and support extended by each individual mentioned above. We acknowledge and appreciate your contributions, which have enriched our learning experience and propelled us towards excellence.

Date:

Place: Kolhapur

INDEX

ABSTRACT

"DeepScan Pro" is an advanced system designed to detect deepfake media, including images, videos, and audio, using state-of-the-art machine learning and deep learning algorithms. It analyzes subtle inconsistencies in media content by leveraging neural networks trained on real and deepfake datasets. The system adopts a multi-modal approach, allowing it to detect anomalies in images, videos, and audio by examining spatial features, motion patterns, and waveform inconsistencies. By incorporating advanced preprocessing techniques such as feature extraction, frame analysis, and spectral analysis, it ensures high accuracy and efficiency in deepfake detection.

As the misuse of deepfake technology continues to pose significant risks in areas like cybersecurity, media integrity, and public trust, "DeepScan Pro" serves as a crucial defense mechanism. It has demonstrated high effectiveness in preliminary tests across diverse datasets, showcasing its ability to identify manipulated content with precision. The system's real-time processing capabilities make it highly adaptable for applications in digital forensics, media verification, and law enforcement.

Designed for scalability, "DeepScan Pro" integrates an intuitive user-friendly interface, ensuring accessibility for users across various industries, including journalism, government agencies, and financial institutions. As deepfake technology evolves, the system continually updates its models with new data and emerging detection techniques to stay ahead of sophisticated manipulation methods. With its robust architecture, real-time analysis, and continual improvement, "DeepScan Pro" establishes itself as a powerful tool in the fight against deepfake deception, safeguarding the authenticity of digital media.

INTRODUCTION

The advent of deepfake technology has revolutionized digital media, enabling the creation of highly convincing but entirely fake content, including images, videos, and audio. While this technology has fascinating applications, it also poses significant risks, particularly in spreading misinformation, committing fraud, and violating personal privacy. As deepfakes become increasingly sophisticated, the need for reliable detection methods has become more pressing than ever.

"DeepScan Pro" is a system developed to address this challenge by providing an advanced solution for detecting deepfake media. The system leverages state-of-the-art machine learning and deep learning algorithms to analyze and identify manipulated content. By integrating cutting-edge neural networks with comprehensive preprocessing techniques, "DeepScan Pro" ensures high accuracy and reliability in its detection process. Designed to handle various types of media, the system performs detailed analysis at multiple levels. Whether it's features examination of images, frame-by-frame scrutiny of videos, or waveform analysis of audio. This multi-faceted approach enables "DeepScan Pro" to effectively identify deepfakes, making it a crucial tool in the ongoing effort to protect the integrity of digital media.

As deepfake technology continues to evolve, the complexity of detecting such manipulations increases, making traditional methods less effective. "DeepScan Pro" addresses this by not only analyzing individual media elements but also considering contextual factors like lighting anomalies, unnatural facial movements, and irregularities in audio patterns. This comprehensive analysis enables the system to stay ahead of evolving deepfake techniques. With real-time processing capabilities, "DeepScan Pro" is an invaluable tool for industries such as law enforcement, media, and social platforms where rapid identification of manipulated content is essential to maintaining trust and authenticity in digital media.

- **Need of the Work:**

The rapid advancement of artificial intelligence has enabled the creation of deepfake media, which can manipulate images, videos, and audio with near-perfect realism. While deepfake technology has legitimate applications in entertainment and content creation, its misuse poses serious risks, including misinformation, identity theft, financial fraud, and threats to national security. The ability to fabricate convincing false media has led to growing concerns in areas such as journalism, law enforcement, and cybersecurity.

Detecting deepfake content is crucial for maintaining trust and authenticity in digital communications. Traditional detection methods rely on manual verification and watermarking techniques, which are no longer sufficient against AI-generated forgeries. Therefore, an automated, AI-driven solution capable of analyzing and identifying deepfakes in real time is essential. "DeepScan Pro" addresses this need by providing an efficient and scalable system for detecting manipulated media, ensuring digital integrity, and mitigating the potential harms associated with deepfake technology.

- **Existing Systems:**

Several deepfake detection systems have been developed to identify and counter AI-generated fake media. Deepware Scanner is a cloud-based tool that analyzes videos for manipulation by detecting inconsistencies in facial expressions, lighting, and compression artifacts. While widely used for quick verification, it is limited to video analysis and does not support image or audio deepfake detection. Sensity AI is a real-time detection platform designed for businesses and government agencies, utilizing deep neural networks to analyze manipulated images, videos, and synthetic profile pictures. It provides detailed reports on digital forgeries but is restricted by commercial licensing, limiting its accessibility for the general public. Microsoft Video Authenticator is another deepfake detection tool that evaluates video authenticity in real-time by analyzing frame-by-frame distortions and blending artifacts. However, its access is restricted to select organizations, making it unavailable for widespread public use.

While these systems offer valuable deepfake detection capabilities, they have notable limitations. Most focus only on image and video analysis, neglecting deepfake audio detection, which is becoming an increasingly critical threat. Additionally, scalability remains a challenge, as many existing models struggle with real-time processing and large datasets. Furthermore, limited public access to some of these tools restricts their widespread adoption. "DeepScan Pro" aims to overcome these challenges by integrating multi-modal deepfake detection, covering images, videos, and audio with real-time processing and improved accessibility, making it a more robust and comprehensive solution.

- **Proposed System:**

"DeepScan Pro" introduces a multi-modal deepfake detection system that improves upon existing methods by integrating machine learning and deep learning techniques for enhanced accuracy and reliability. The key features of the proposed system include:

- **Multi-Modal Analysis** – Unlike traditional detection methods, "DeepScan Pro" simultaneously examines images, videos, and audio for deepfake detection, ensuring comprehensive media verification.

- **Advanced Neural Networks –** The system employs deep learning models, including SwinTransformer for image analysis and CNN-based architectures for audio waveform analysis. These models identify deepfake artifacts such as unnatural facial expressions, lighting anomalies, and voice irregularities.
- **Robust Preprocessing Techniques –** The system applies feature extraction, frame analysis, and spectral analysis to improve detection accuracy by focusing on critical patterns in manipulated media.
- **Real-Time Processing –** The system is optimized for fast and efficient deepfake detection, making it suitable for applications in journalism, cybersecurity, and law enforcement.
- **User-Friendly Interface –** A Flet-based UI ensures easy accessibility for users, allowing them to upload and analyze media effortlessly.

By leveraging these advancements, "DeepScan Pro" offers a scalable, accurate, and real-time deepfake detection solution, addressing the limitations of existing systems while enhancing digital security and trustworthiness.

LITERATURE SURVEY

Yan Wang, Qindong Sun, Dongzhu Rong, and Rong Geng [1] developed a deepfake video detection model designed specifically for compressed videos. Their approach integrates frequency domain adaptive notch filtering, spatial residual denoising, and attention-based feature fusion, effectively minimizing compression artifacts that often degrade detection accuracy. However, the model is less effective on uncompressed or lightly compressed videos, making it less adaptable to different types of manipulated media.

Jayashre K and Amsaprabhaa M [2] introduced the HODFF-DD framework, which combines Inception ResNet and BiLSTM networks to detect spatial and temporal inconsistencies in deepfake videos. The model optimizes feature extraction and temporal analysis using the spotted hyena optimization algorithm, enhancing performance across datasets. While highly effective, it was evaluated only on single-face videos, limiting its suitability for multi-face deepfake scenarios.

Rahul TP, PR Aravind, Ranjith C, Usamath Nechiyil, and Nandakumar Paramparambath [3] focused on audio deepfake detection, utilizing ResNet-34 and transfer learning. The system extracts Mel-spectrogram features to identify synthetic speech patterns in Automatic Speaker Verification (ASV) systems. While transfer learning improves detection performance, the model struggles with real-world testing, indicating the need for further refinement.

Rul Liu, Jinhua Zhang, and Guanglai Gao [4] proposed the MSCR-ADD framework, a novel audio deepfake detection system that transforms mono audio signals into binaural representations. By leveraging multi-space channel representations, this method enhances detection accuracy by capturing channel-specific and channel-invariant features. However, its ability to detect deepfake singing voices remains uncertain, highlighting an area for further development.

Preeti Sharma, Manoj Kumar, and Hitesh Kumar Sharma [5] introduced a GAN-CNN ensemble model for detecting social media deepfakes. Their approach combines generative adversarial networks (GANs) and convolutional neural networks (CNNs), using generative replay to prevent catastrophic forgetting when training on evolving deepfake techniques. While effective, the model lacks a detailed discussion on its limitations and adaptability to new deepfake manipulations.

Bo Wang, Xiaohan Wu, Fei Wang, Yushu Zhang, Fei Wei, and Zengren Song [6] developed a Spatial-Frequency Fusion Branch (SFFB) method, which integrates spatial and frequency domain features within a knowledge distillation framework. This approach significantly enhances detection accuracy on datasets such as FaceForensics++ and Celeb-DF, which often contain compression artifacts. However, its accuracy declines with highly compressed datasets, indicating a need for further enhancements.

Chen Li, Haoran Wang, and Xiang Gao [7] proposed a multi-modal deepfake detection system that integrates image, video, and audio analysis using a Transformer-based architecture. By extracting cross-modal dependencies, the model enhances deepfake detection across different types of manipulated content. Unlike single-modal systems that analyze only one media type, this approach improves accuracy by detecting shared deepfake artifacts across multiple modalities. However, its high computational cost makes real-time processing difficult, requiring future optimizations to improve efficiency.

Recent deepfake detection methods target specific challenges across media types. Yan Wang et al. focused on compressed videos using frequency filtering and attention fusion but lacked adaptability to uncompressed formats. Jayashre K and Amsaprabhaa M's HODFF-DD framework combined Inception ResNet and BiLSTM for spatial-temporal detection, though limited to single-face videos. Rahul TP et al. used ResNet-34 and transfer learning for audio detection, performing well in controlled settings but struggling in real-world use. Rul Liu et al. enhanced audio detection with binaural representations, though deepfake singing detection remains untested. Preeti Sharma et al. combined GANs and CNNs with generative replay for evolving deepfakes but didn’t fully address limitations. Bo Wang et al. integrated spatial and frequency features in a knowledge distillation framework, effective on common datasets but less so on heavily compressed ones. Chen Li et al. proposed a multi-modal Transformer system for cross-media detection, offering high accuracy but limited by computational demands.

Motivation:

The motivation behind developing DeepScan Pro came from our awareness of the growing threat that deepfakes pose to society. As AI-generated media becomes increasingly realistic and accessible, we saw the potential for misuse in spreading misinformation, manipulating public opinion, and compromising digital trust. This concern inspired us to take action by creating a comprehensive deepfake detection system capable of analyzing images, audio, and video. Through this project, we aimed to apply our knowledge of machine learning and media forensics to build a practical solution that helps identify synthetic content and supports the fight for media authenticity and digital integrity.

PROBLEM STATEMENT & OBJECTIVES

Problem Statement:

"To develop a system that detect deepfakes in media such as images, videos, and audio."

Objectives:

1. To develop a multi-modal deepfake detection system integrating images, videos, and audio.
2. To design and implement an improved user interface for enhanced user experience.
3. To test and validate the system on a comprehensive dataset of diverse deepfake media, ensuring its applicability in real-world scenarios.

REQUIREMENT ANALYSIS

a. Dataset Description:

To train, test, and validate the DeepScan Pro system for accurate deepfake detection, we utilized a variety of datasets across images, audio, and video domains. These datasets include both real and AI-generated media, ensuring that the model generalizes well to real-world manipulations. The datasets were curated from both public sources and custom-created to align with the Indian linguistic and visual context, making the system regionally relevant and globally effective.

1. Image Dataset:

A total of 13,000 face images were used, with an equal split between real and AI-generated images.

1. **Custom Dataset:** We generated 1,500 real human facial images using camera devices and gathered 1,500 synthetic facial images created using AI-based face generators.
2. **Public Face Datasets:** Additional images were incorporated from existing facial datasets to include varied age groups, ethnicities, lighting conditions, and image resolutions. These help the system detect inconsistencies introduced by AI-generated facial manipulations.

2. Audio Dataset:

For detecting deepfake audio, especially in the Indian context, a combination of real-world and synthetic voice datasets was employed.

1. **Indian Voice Dataset (Custom):** Real and fake audio clips were recorded and synthesized in various Indian languages and accents using Text-to-Speech (TTS) systems.
2. **ASVspoof Dataset:** A benchmark dataset used for detecting spoofing attacks in speaker verification systems, it includes diverse spoofing techniques such as voice conversion and replay attacks.
3. **SceneFake Dataset:** This dataset includes real and deepfake audio from real-world environments, enhancing the system’s performance on conversational and background-noise-affected audio clips.

3. **Video Dataset:**

   Video data is central to deepfake detection as it requires spatial-temporal feature extraction.

   1. **Celeb-DF V2:** Offers high-resolution deepfakes of celebrities with minimal artifacts, making it valuable for testing system performance on subtle manipulations.
   2. **Hindi Audio-Video Dataset (Custom):** A locally curated dataset of Indian-origin videos combining facial and vocal manipulations in the Hindi language, designed to test the multi-modal capabilities of DeepScan Pro in regional contexts.

b. Hardware Requirements:

Given the compute-intensive nature of deep learning tasks, especially for image and audio processing, a reasonably powerful hardware setup was required for development, training, and testing of the system. The following configuration was used and recommended:

1. **Operating System:** Compatible with Windows 10/11, macOS, Ubuntu, or other Linux distributions to allow flexibility across different platforms.
2. **Processor:**
   1. Minimum: Intel Core i5 (8th Gen or above) or AMD Ryzen 5.
   2. Recommended: Intel Core i7/i9 or AMD Ryzen 7/9 for faster training times and multi-threaded operations.
3. **RAM:**
   1. Minimum: 8 GB
   2. Recommended: 16 GB or higher for smooth model training and dataset handling.
4. **Storage:**
   1. Minimum 512 GB SSD or HDD for storing large datasets and model weights.
   2. Recommended 1 TB SSD for better read/write speed during model training.
5. **Graphics Card (GPU):**
   1. Optional for basic use.
   2. Recommended: NVIDIA GTX 1660 / RTX 2060 or higher (with CUDA support) for GPU-accelerated training.

c. Software Requirements:

To support the diverse functionalities of DeepScan Pro - including machine learning training, media preprocessing, and real-time detection—various software environments and platforms were utilized:

1. **Programming Language:**
    1. **Python 3.8 or later:** Chosen for its readability, large community, and rich ecosystem of machine learning and data science libraries.
2. **Development Environments:**
    1. **Google Colab:** Used for training and testing models on GPU-enabled environments.
    2. **Jupyter Notebook:** For prototyping and visualization of datasets and training metrics.
    3. **Visual Studio Code (VS Code):** For code development, debugging, and version control.
3. **Operating System Compatibility:**

    Cross-platform support for Windows, Linux (Ubuntu), and macOS systems, ensuring deployment flexibility.
4. **Version Control System:**

    Git and **GitHub** for collaborative code management, version tracking, and documentation.

d. Libraries Used:

The system integrates a wide range of Python libraries for media preprocessing, machine learning model development, feature extraction, visualization, and UI deployment.

- **Machine Learning & Deep Learning Libraries:**
    1. **TensorFlow:** For training audio-based deepfake detection models and performing classification tasks.
    2. **PyTorch:** Core framework for training CNN, ResNet, and custom models for image, audio, and video classification.
    3. **Scikit-learn:** For Splitting data and evaluating metrics like accuracy, F1-score, and confusion matrix.
    4. **Transformers:** From HuggingFace, used to integrate powerful pretrained models like Swin Transformer (image) and Wav2Vec2 (audio) for feature extraction.
- **Image & Video Processing Libraries:**
    1. **OpenCV:** For video frame extraction, resizing, face detection, and manipulation.

2. **PIL (Pillow):** For image processing tasks such as cropping, resizing, and format conversion.
3. **DeepFace:** For facial recognition, feature alignment, and comparison.
4. **face_recognition:** Lightweight library used for detecting and comparing human faces in images.

- **Audio Processing Libraries:**
  1. **Librosa:** For extracting audio features such as MFCCs, Mel-spectrograms, ZCR, and chroma.
  2. **Soundfile:** For loading/saving .wav and other audio formats.
  3. **Wav2Vec2:** Used for extracting embeddings from raw speech signals for deepfake audio detection.

- **Data Handling & Visualization:**
  1. **NumPy:** For handling matrix computations and tensor operations.
  2. **Pandas:** For data cleaning, dataset labeling, and manipulation.
  3. **Matplotlib & Seaborn:** For visualizing model training, evaluation results, confusion matrices, and data distributions.
  4. **Captum:** For interpreting PyTorch models, especially for visualizing attention weights and feature importance.

- **Utility & Backend Integration:**
  1. **tempfile:** Used to store temporary media files uploaded by the user.
  2. **uuid:** For generating unique filenames and session IDs during uploads or database entries.
  3. **math:** Used for mathematical calculations in preprocessing and metrics.
  4. **webbrowser:** To launch browser sessions for UI access or result visualization.
  5. **pyrebase:** For managing Firebase authentication and real-time database for user management and media logging.

- **User Interface Development:**
  1. **Flet:** A modern cross-platform UI library used to develop a lightweight, interactive interface allowing users to signup, login, upload media and view deepfake detection results in real time.

SYSTEM DESIGN & IMPLEMENTATION

a. System Architecture:

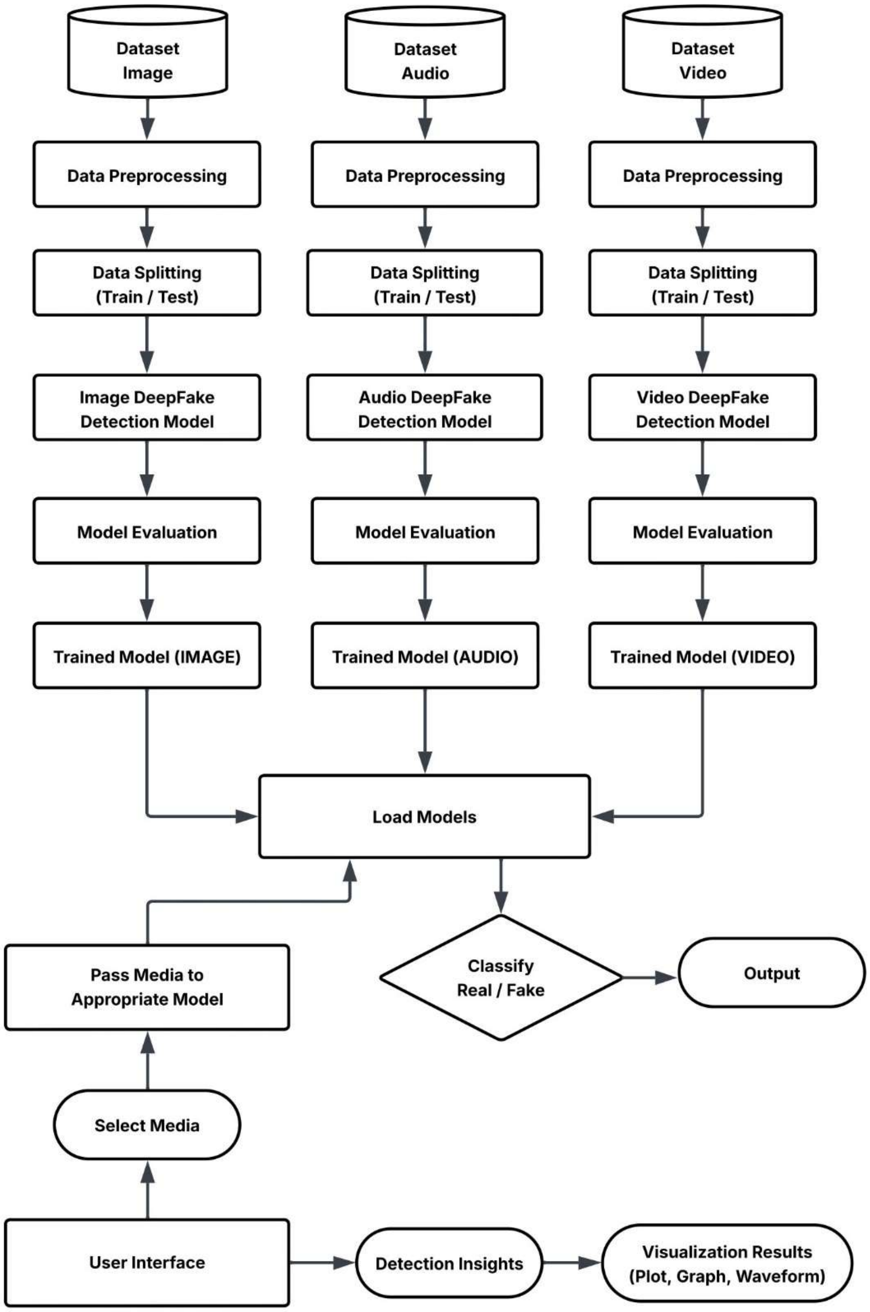


Fig a. System Architecture of DeepScan Pro

- **Data Collection:**

1. The DeepScan Pro system is developed using a diverse and carefully curated collection of custom datasets that include images, videos, and audio clips. These datasets are sourced from various domains such as a custom image dataset, human faces dataset,

ASVspoof 2019, SceneFake (both audio and visual), a Hindi custom audio dataset, CelebDF v2, and a Hindi audio-video dataset.

2. By incorporating both authentic and manipulated media across languages and formats, the system effectively learns to identify subtle visual and auditory cues that differentiate real content from deepfakes.

- **Data Preprocessing:**
  1. Image Preprocessing: Facial regions are detected and cropped using face_recognition and DeepFace, followed by resizing and feature extraction via Swin Transformer.
  2. Audio Preprocessing: Audio signals are processed using Librosa, converted to embeddings with Wav2Vec2, and optionally analyzed with MFCCs or Mel-spectrograms.
  3. Video Preprocessing: Frames are extracted using OpenCV, resized, and passed through a ResNeXt50 CNN for spatial features, followed by LSTM to capture temporal patterns.
- **Model Training:**
  1. **Image – Model Training:** Swin Transformer is used to extract features from facial images, which are then classified using a custom PyTorch-based neural network. The classifier includes normalization, dense layers, and non-linear activations to capture manipulation cues. It is trained on custom and public face datasets to distinguish real from fake.
  2. **Audio – Model Training:** Wav2Vec2 embeddings are extracted from audio samples and fed into a BiLSTM with attention to learn speech patterns. This helps focus on subtle voice manipulations often present in fake audios. The final classifier predicts authenticity with high accuracy.
  3. **Video – Model Training:** Each video frame is processed using ResNeXt50 for spatial features, then an LSTM captures temporal inconsistencies. The model learns how deepfakes evolve across frames and outputs a real or fake label. Training is done on short clips from multiple datasets.
- **Model Evaluation:**
  1. Model evaluation in DeepScan Pro uses balanced real and fake samples from all media types. Metrics like accuracy, precision, and F1-score help measure detection performance. This ensures the model is reliable, unbiased, and effective across various deepfake formats.

- **Interface Development:**
  1. A user-friendly interface is developed using Flet.
  2. The interface allows users to input various types of media (images, videos, or audio) for analysis.
  3. It provides real-time feedback on whether the media is authentic or manipulated and visualizes the areas where inconsistencies are detected.

b. Data Flow Diagram (DFD)

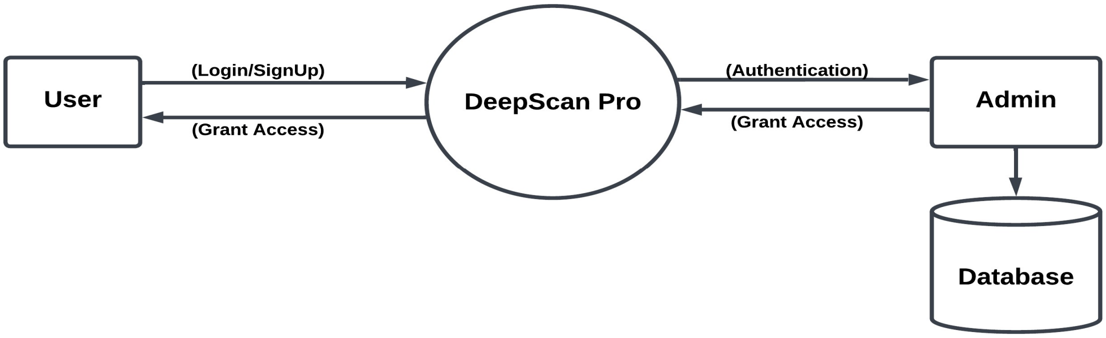


Fig b. 1. Data Flow Diagram Level - 0

This Level-0 DFD provides a simplified overview of how data flows through a DeepScan Pro System, starting user login, being authenticated by the admin, and resulting in a granting access to the user. This high-level diagram does not go into the details of the processes within the DeepScan Pro or the specifics of the data.

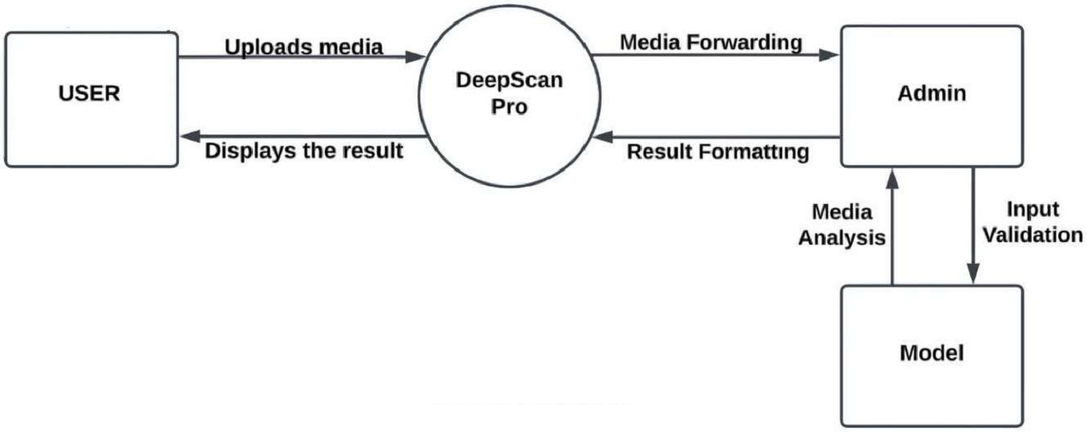


Fig b. 2. Data Flow Diagram Level - 1

This Level-1 DFD breaks down the initial high-level view (Level-0) into more specific steps, showing how input media is first pre-processed, validated and analyzed and being fed to a machine learning model to generate predictions. This diagram provides a clearer understanding of the intermediate steps involved in the data flow process.

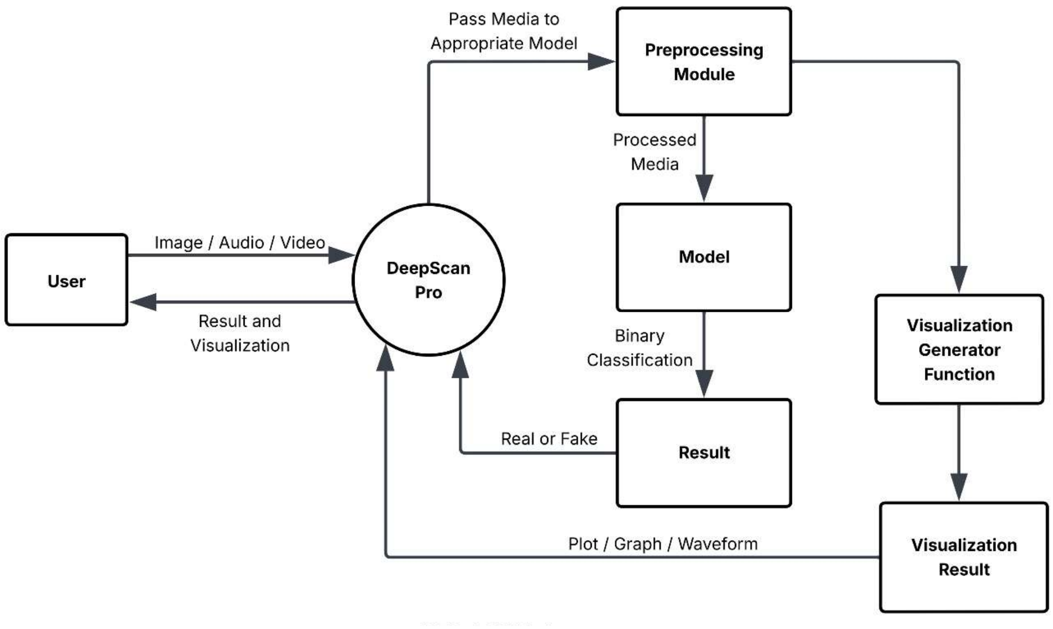


Fig b. 3. Data Flow Diagram Level - 2

This Level-2 DFD breaks down the process further into more detailed steps, showing how data is pre-processed, features are extracted, the model is selected and applied, and how predictions are generated, categorized and visualized. This detailed view helps in understanding the intricate workings and dependencies within the system.

c. Use Case Diagram

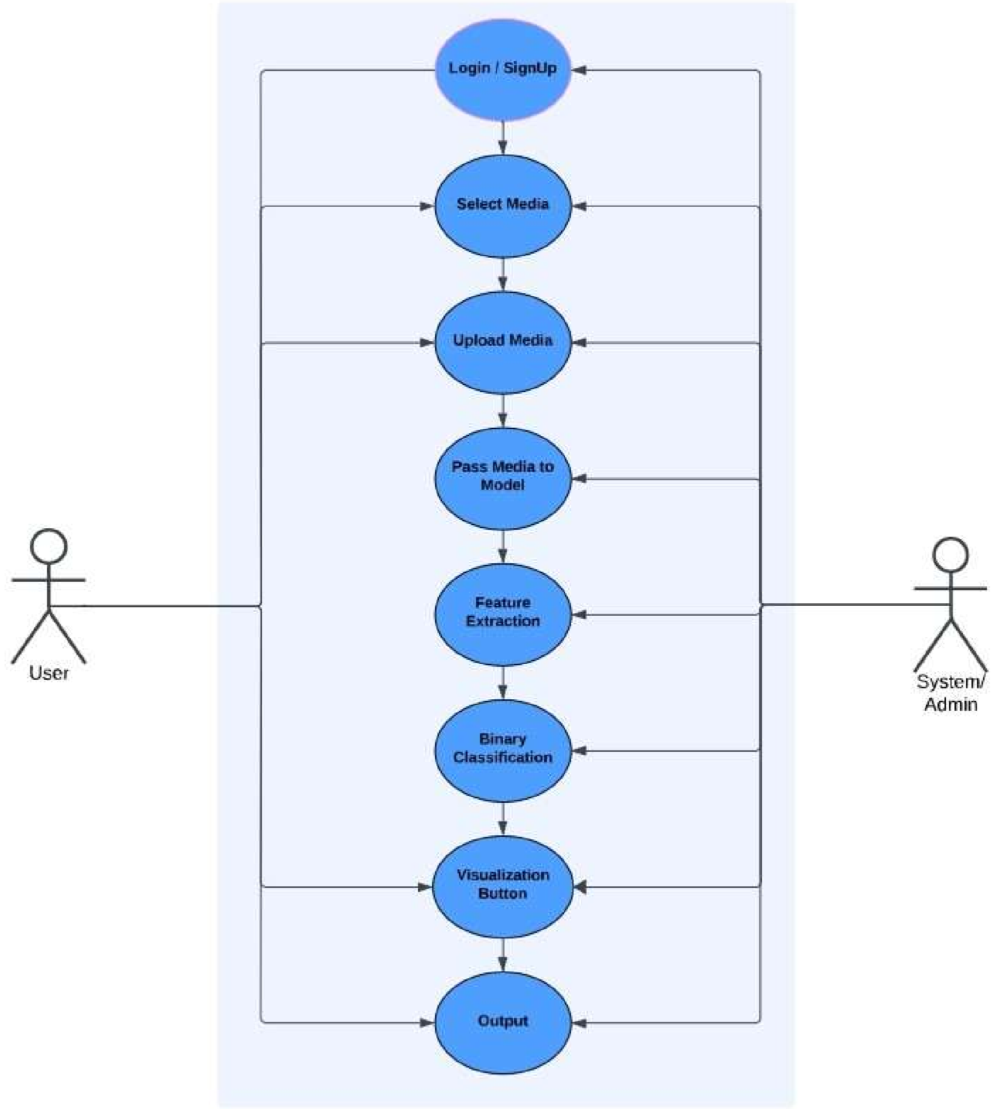


Fig c. Use Case Diagram

The use case diagram provides a high-level overview of the different actions that the user can take when interacting with the DeepScan Pro System, as well as the actions that the system can take to generate predictions.

d. Implementation Details:

The implementation of DeepScan Pro involves a modular and multi-modal architecture designed to detect deepfake content in images, videos, and audio using advanced machine learning and deep learning algorithms. The system integrates several key components such as data preprocessing, model training, prediction pipelines, and real-time user interaction. Below are the core aspects of its implementation:

Algorithm:

Input: User credentials, uploaded media file

Output: Classification result (Real or Fake)

```
begin
    1. Display start page with options: [Signup], [Login]
    2. If user selects [Signup]:
        a) Collect username, email, and password
        b) Register user using Firebase Authentication
        c) Save user information to Firebase Database
    3. Else if user selects [Login]: Collect user credentials (email, password) &
       Authenticate using Firebase.
            If login is successful, redirect to [Home Page]
            Else, show error message and allow retry
    4. If user selects "Upload Media" on [Home Page]:
        a) Open file picker for media selection
        b) Save uploaded media file temporarily
        c) Display selected file on UI
    5. On "Detect", button click:
       a) Identify file type (image / audio / video) using file extension.
    a) If media type is Image:
            i.   Apply image preprocessing (resize, normalize, face detection)
            ii.  Extract features using Swin Transformer
            iii. Predict using image classification model
    b) Else if media type is Audio:
            i.   Convert audio to Mel-spectrogram / extract Wav2Vec2 embeddings
            ii.  Predict using BiLSTM + Attention-based model
    c) Else if media type is Video:
            i.   Extract frames using OpenCV, Crop faces and preprocess frames
            ii.   Extract features using ResNeXt
            iii.   Pass sequence to LSTM model for temporal classification
    6. Collect output from model:
            a) Generate prediction probability
            b) Classify as Real or Fake based on threshold
            c) Display result on UI with label and confidence score
    7. On “Logout”, button click: On successful logout, redirect to [Login Page]
end
```

RESULT ANALYSIS

- **Image Classification:**

Three models were evaluated for image classification: ResNet18, EfficientNetB4, and Swin Transformer. ResNet18 achieved 67% accuracy, while EfficientNetB4 improved it to 75%. Swin Transformer output formed both with 90% accuracy and strong precision, recall, and F1 scores. Hence, Swin was selected for its superior and consistent performance.

Model	Accuracy_score	Precision_score	Recall_score	F1_score
ResNet18	0.670	0.68	0.65	0.66
EfficientNetB4	0.750	0.76	0.73	0.74
SwinTransformer	0.900	0.91	0.89	0.90

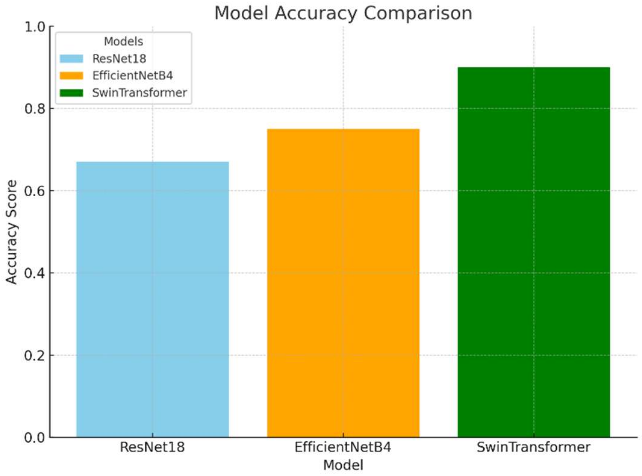


Fig. a. Image Model Accuracy Comparison

- **Audio Classification:**

Audio CNN achieved 64% accuracy, while Wav2Vec2 boosted it to 76%. Wav2Vec2 leveraged pretrained audio representations for better performance. It showed improved generalization and feature extraction over CNN. Hence, Wav2Vec2 was chosen for audio-based tasks.

Model	Accuracy_score	Precision_score	Recall_score	F1_score
Audio CNN	0.640	0.65	0.62	0.63
Wav2Vec2	0.760	0.78	0.75	0.76

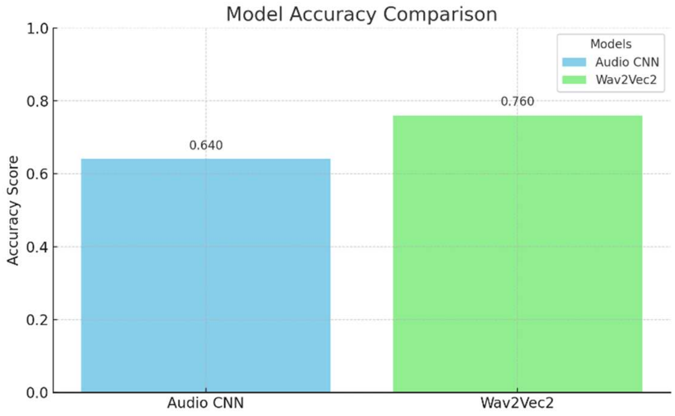


Fig. b. Audio Model Accuracy Comparison

- **Video Classification:**

A hybrid model using ResNeXt and LSTM achieved 87% accuracy on video data. ResNeXt handled spatial features while LSTM captured temporal dynamics. This combination proved effective for deepfake video detection. Thus, ResNeXt + LSTM was selected as the video classification model.

Model	**Accuracy_score**	**Precision_score**	**Recall_score**	**F1_score**
ResNeXt + LSTM	0.870	0.88	0.86	0.87

Comparison Criteria	**Deepware Scanner**	**DeepScan Pro**
Algorithm Complexity	Basic CNN-based analysis of facial expressions and lighting	Advanced models: Swin Transformer (Image), BiLSTM (Audio), ResNeXt50 + LSTM (Video)
Model Accuracy (Image)	Not applicable (video-only)	ResNet18 → EfficientNetB4 → Swin Transformer with up to 90% accuracy
Audio Deepfake Detection	Not supported	Implemented with CNN → BiLSTM + Attention (76% accuracy)

Real-time Processing	Partial – cloud-based, not optimized for real-time	Fully optimized for real-time local or cloud processing
Dataset Adaptability	Limited; no regional/language-specific datasets	Uses custom datasets including Indian languages and local faces
User Accessibility	Limited to specific platforms, no public release	Open-access with intuitive Flet-based UI, easy for end-users

Media Type	DeepScan Pro (%)	Deepware Scanner (%)
Image	90	Not Available
Audio	76	Not Available
Video	97	70

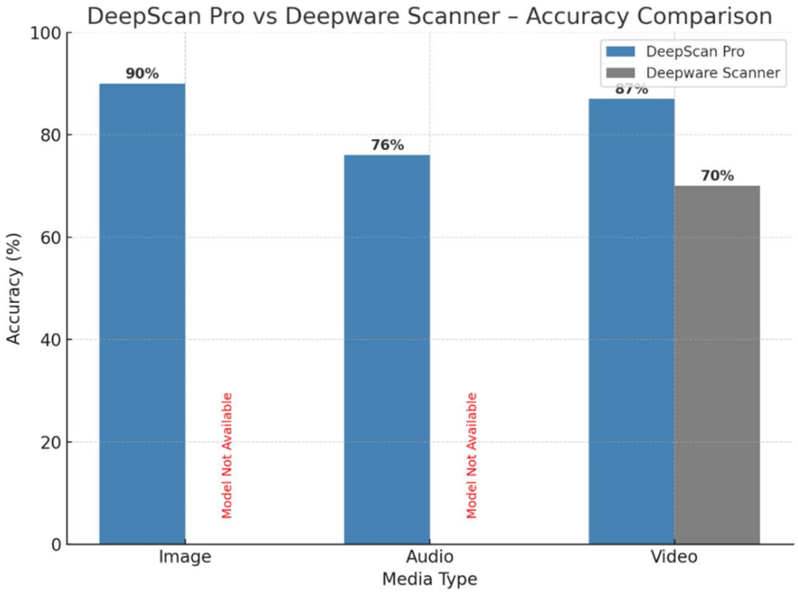


Fig. DeepScan Pro vs Deepware Scanner

CONCLUSION & FUTURE SCOPE

Conclusion:

DeepScan Pro has proven to be a robust, multi-modal deepfake detection system capable of analyzing and identifying manipulated media across images, audio, and video. By utilizing advanced deep learning models like Swin Transformer, Wav2Vec2 with BiLSTM, and ResNeXt50 with LSTM, the system effectively captures spatial, spectral, and temporal anomalies that are characteristic of deepfakes. The integration of a user-friendly interface and real-time analysis features ensures that the system is practical, efficient, and suitable for real-world applications, particularly in areas like journalism, digital forensics, and law enforcement.

In comparison with existing systems, DeepScan Pro addresses key limitations such as limited modality coverage, high hardware dependencies, and restricted access. With its diverse dataset strategy—including region-specific content—it delivers higher accuracy and contextual relevance. The system's modular architecture and explainable outputs further enhance trust and transparency, making it a reliable solution for detecting deepfake threats in evolving digital landscapes.

Future Scope:

As deepfake generation techniques continue to advance, DeepScan Pro can evolve by incorporating more advanced transformer-based models and self-supervised learning techniques for better generalization. One key area for improvement is expanding image-based detection beyond facial manipulations to include full-image deepfakes such as synthetic backgrounds, object tampering, or altered contexts—enhancing the system's ability to identify broader forms of visual misinformation. Future versions may also include multilingual audio processing, integration with real-time social media monitoring tools, and deployment as a browser extension or mobile app to increase reach and usability. Additionally, enhancing the system's interpretability and adapting it for low-resource environments will further strengthen its applicability in diverse global settings.

REFERENCES

[1] Wang, Y., Sun, Q., Rong, D., & Geng, R., "Multi-domain awareness for compressed deepfake videos detection over social networks guided by common mechanisms between artifacts," Computer Vision and Image Understanding, Vol. 104, No. 2, pp. 72-81, 2024. ScienceDirect.

[2] Jayashre, K., & Amsaprabhaa, M., "Safeguarding media integrity: A hybrid optimized deep feature fusion based deepfake detection in videos," Computers & Security, Vol. 103, No. 4, pp. 860-868, 2024. ScienceDirect.

[3] Rahul, T.P., Aravind, P.R., Ranjith, C., Nechiyil, U., & Paramparambath, N., "Audio spoofing verification using deep convolutional neural networks by transfer learning," arXiv preprint arXiv:2008.03464, Vol. 28, No. 2, pp. 34-42, 2024. ScienceDirect.

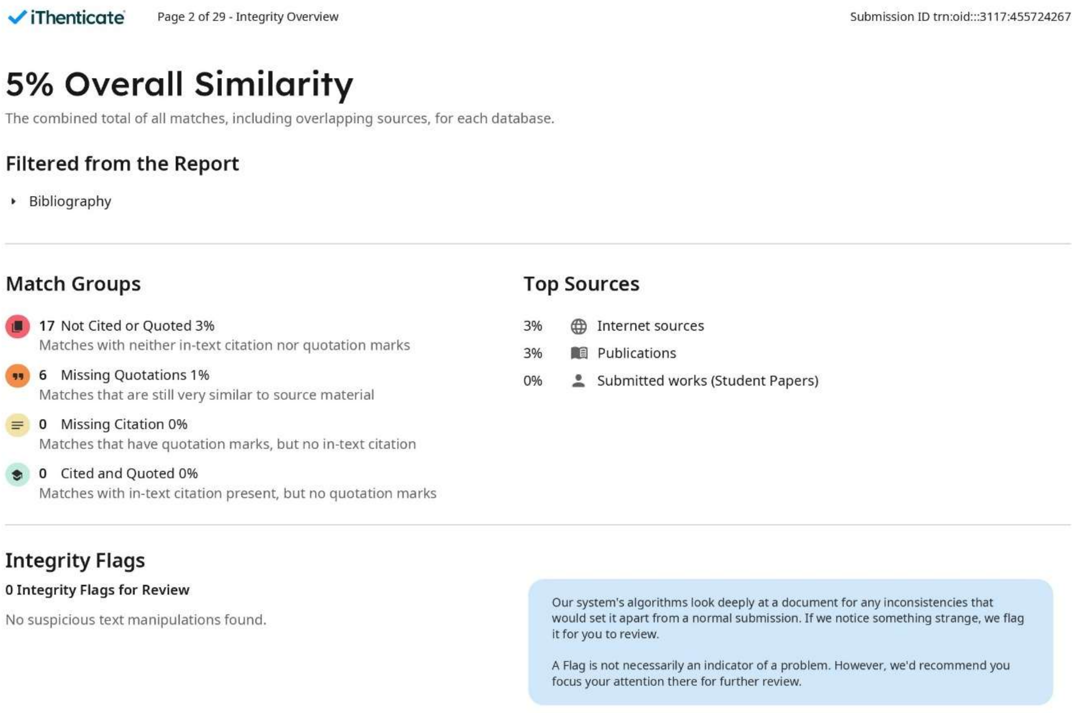
[4] Liu, R., Zhang, J., & Gao, G., "Multi-space channel representation learning for mono-to-binaural conversion based audio deepfake detection," Information Fusion, Vol. 102, No. 2, pp. 257-264, 2024. ScienceDirect.

[5] Sharma, P., Kumar, M., & Sharma, H.K., "GAN-CNN Ensemble: A robust deepfake detection model of social media images using minimized catastrophic forgetting and generative replay technique," Procedia Computer Science, Vol. 204, No. 5, pp. 90-99, 2024. ScienceDirect.

[6] Wang, B., Wu, X., Wang, F., Zhang, Y., Wei, F., & Song, Z., "Spatial-frequency feature fusion based deepfake detection through knowledge distillation," Journal of Information Fusion, Vol. 65, No. 1, pp. 257-268, 2024. ScienceDirect.

[7] Li, C., Wang, H., & Gao, X. (2023). Multi-modal deepfake detection using Transformer-based cross-modal fusion. Journal of Multimedia Forensics and Security, 58(2), 189–202. Elsevier.

PLAGIARISM REPORT



iThenticate Page 2 of 29 - Integrity Overview Submission ID trn:oid:::3117:455724267

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

- **17** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **6** Missing Quotations 1%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3%	Internet sources
3%	Publications
0%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

PUBLICATION DETAILS

We have published the research paper entitled **“DeepScan Pro - A Deepfake Detector”** in the **IJCRT (International Journal of Creative Research Thoughts)**, Volume: 13, Issue: 5, May 2025, ISSN: 2320-2882, Paper ID: 285218, Page Number(s): b673-b680. The journal has an **Impact Factor of 7.97** (calculated by Google Scholar and Semantic Scholar).