TruthCheck: AI-Based Fake Media Detection System

M.Akila
Assistant Professor
Department of IT
K.L.N. College of Engineering,
Sivagangai, India

V.K.Poornimaa Shri
UG Scholar
Department of IT
K.L.N. College of Engineering,
Sivagangai, India

V.Valentina
UG Scholar
Department of IT
K.L.N. College of Engineering,
Sivagangai, India

***Abstract* - The rapid growth of digital platforms has led to an increase in the spread of fake and manipulated media content across various formats such as text, images, videos, and audio. Identifying such misleading information has become a challenging task due to the advancement of artificial intelligence techniques that generate highly realistic fake content.This paper presents TruthCheck, a multi-modal fake media detection system designed to analyze and classify different types of digital content within a unified framework. The system utilizes machine learning and deep learning models to process input data through stages such as preprocessing, feature extraction, and classification. Each media type is handled using dedicated modules to ensure effective detection.The proposed system provides outputs in the form of truth score, fake score, and confidence level, improving transparency and user understanding. The use of lightweight models ensures efficient performance on standard computing environments. The results demonstrate that the system is capable of accurately detecting fake content across multiple media formats, making it suitable for real-world applications.**

Keywords: Fake News Detection, Deepfake Detection, Artificial Intelligence, Explainable AI, Multimedia Analysis

I. INTRODUCTION

In recent years, the increasing reliance on digital platforms for information sharing has created new challenges in verifying the authenticity of content. Online sources such as social media, news portals, and multimedia platforms enable rapid distribution of information, but they also contribute to the uncontrolled spread of misleading and manipulated content.

Advancements in artificial intelligence have further intensified this issue by enabling the creation of highly realistic synthetic media. Techniques such as deep learning-based content generation allow the production of altered images, fabricated videos, synthetic speech, and misleading textual information. These forms of manipulated content are often difficult to identify using traditional verification methods, making users vulnerable to misinformation.

Most existing detection approaches are designed to handle only a specific type of data, such as text-based fake news or image-based deepfake detection. This fragmented approach limits their effectiveness in real-world scenarios where multiple forms of media coexist. Additionally, many systems rely on complex models that demand significant computational resources, reducing their accessibility for general users **[2], [6].**

Another important concern is the limited interpretability of current detection systems. Many tools provide only a final classification result without offering insights into how the decision was made. This lack of clarity reduces user confidence and restricts the practical usability of such systems **[1], [4].**

Recent research has explored various advanced techniques such as transformer-based models, convolutional neural networks, and multi-modal frameworks for improving detection accuracy. These approaches have shown promising results in identifying manipulated content across different domains, including text, images, videos, and audio **[3].**

To overcome these limitations, this work introduces TruthCheck, a unified system capable of analyzing diverse forms of digital content within a single framework. The proposed solution focuses on efficiency by utilizing optimized machine learning and deep learning models that can operate on standard computing environments **[6], [11].**

Furthermore, the system emphasizes result interpretability by presenting confidence measures and indicative patterns that support the prediction outcome. This approach not only improves detection capability but also enhances user understanding and trust **[1], [16].**

Deepfake speech detection has also gained significant attention, where models analyze audio patterns and spectrogram features to identify synthetic voice content. Techniques such as neural network-based classifiers have been effectively used to distinguish between real and manipulated audio signals **[5].**

Studies focusing on smart grid and energy optimization introduced real-time electricity management frameworks to improve efficiency and reduce operational costs. While technically advanced, these systems were primarily designed for energy optimization rather than holistic residential governance **[11], [13]**.

Data collection scheduling techniques were proposed to optimize IoT-based utility monitoring networks. These methods enhanced system efficiency and reliability but did not emphasize user-level application interfaces for administrative and resident interaction **[15]**.

By combining multi-modal analysis with transparent output representation, the proposed system contributes toward building a more reliable mechanism for identifying manipulated digital content and reducing the impact of misinformation in online environments **[17], [20].**

II. METHODOLOGY

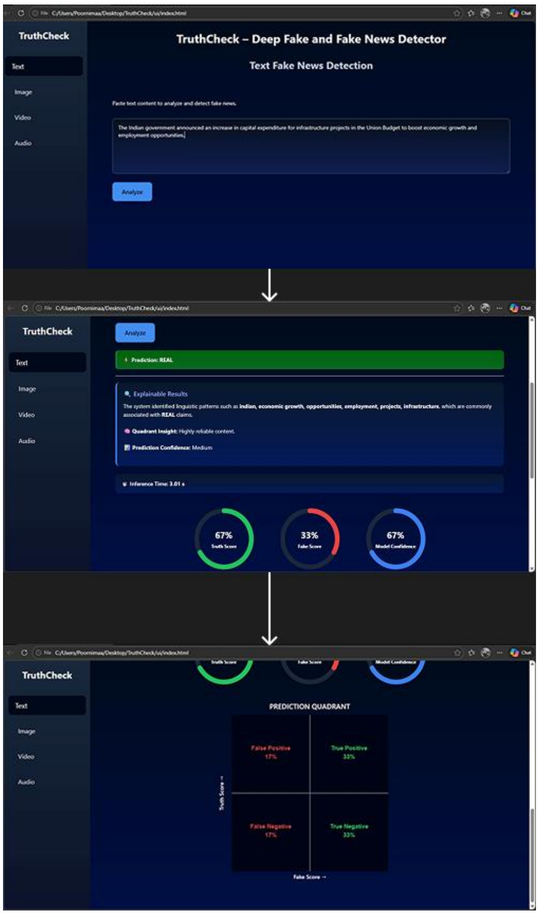
The proposed system, TruthCheck, is developed as a multi-modal fake media detection platform that analyzes various types of digital content including text, images, videos, and audio. The system follows a structured pipeline where user input is first collected and then processed through different stages such as preprocessing, feature extraction, and model prediction.Each type of media is handled by a dedicated module designed specifically to identify patterns of manipulation. The processed data is analyzed using machine learning and deep learning models, which generate prediction results indicating whether the content is real or fake.

The overall workflow ensures efficient handling of multiple media formats within a single system. The results are presented in the form of truth score, fake score, and confidence level to improve transparency and user understanding.

A. Text Analysis Module

The text analysis module is designed to detect fake information from textual data such as news articles, blogs, and social media posts. Initially, the input text undergoes preprocessing steps including removal of punctuation, stop words, and irrelevant characters. This step ensures that only meaningful textual information is retained for further analysis.

After preprocessing, the cleaned text is transformed into numerical format using the TF-IDF (Term Frequency–Inverse Document Frequency) technique. This method helps in identifying the importance of words within the document and across the dataset.


C:/Users/Poornimaa/Desktop/TruthCheck/ui/index.html

TruthCheck – Deep Fake and Fake News Detector

Text Fake News Detection

Paste text content to analyze and detect fake news.

The Indian government announced an increase in capital expenditure for infrastructure projects in the Union Budget to boost economic growth and employment opportunities.

Prediction: REAL

Explainable Results

The system identified linguistic patterns such as **indian, economic growth, opportunities, employment, projects, infrastructure**, which are commonly associated with **REAL** claims.

Quadrant Insight: Highly reliable content.

Prediction Confidence: Medium

Inference Time: 3.01 s

Figure 1: Text Analysis Workflow

A Logistic Regression model is used for classification, as it is efficient and suitable for text-based problems. The model is trained on labeled datasets containing both real and fake news samples. It learns patterns such as word usage, frequency, and contextual differences between genuine and fake content.

Based on the learned features, the model predicts whether the given text is real or fake. The system also provides additional outputs such as truth score and fake score, which indicate the confidence level of the prediction.

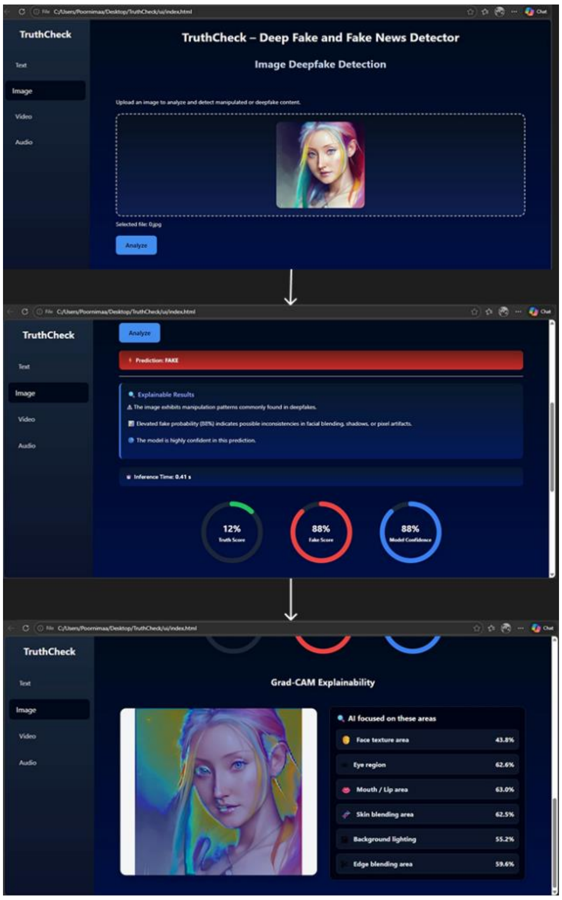
B. Image Analysis Module

he image analysis module is designed to detect manipulated or deepfake images by analyzing visual patterns and identifying inconsistencies present in digital images. With the rapid advancement of deep learning techniques, it has become possible to generate highly realistic fake images, making detection a challenging task. This module aims to address this issue by using robust feature extraction and classification techniques.

Initially, the input image is preprocessed to ensure consistency and compatibility with the model. This includes resizing the image to a fixed resolution, normalizing pixel intensity values, and converting the image into a suitable format for processing. These preprocessing steps help in reducing noise and improving the overall performance of the model.

A Convolutional Neural Network (CNN) is employed for feature extraction, as it is highly effective in capturing spatial information from images. The CNN automatically learns important features such as edges, textures, shapes, and patterns that distinguish real images from manipulated ones. Unlike traditional methods, CNN does not require manual feature engineering, making it more efficient and accurate.

The model focuses on identifying subtle artifacts that are commonly introduced during image manipulation. These include inconsistencies in lighting, unnatural blending of objects, distorted facial structures, and pixel-level irregularities. Such features are often difficult to detect by human observation but can be effectively captured by deep learning models.



TruthCheck – Deep Fake and Fake News Detector

Image Deepfake Detection

Upload an image to analyze and detect manipulated or deepfake content.

Selected file: 0.jpg

Prediction: FAKE

Explainable Results

⚠ The image exhibits manipulation patterns commonly found in deepfakes.

Elevated fake probability (88%) indicates possible inconsistencies in facial blending, shadows, or pixel artifacts.

The model is highly confident in this prediction.

Inference Time: 0.41 s

12% Truth Score

88% Fake Score

88% Model Confidence

Grad-CAM Explainability

AI focused on these areas

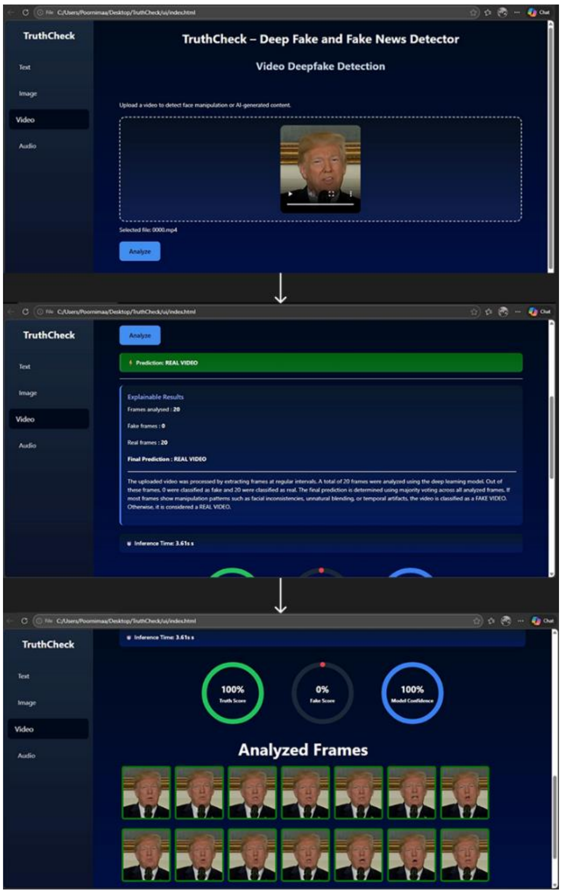
Face texture area	43.8%
Eye region	62.6%
Mouth / Lip area	63.0%
Skin blending area	62.5%
Background lighting	55.2%
Edge blending area	59.6%

Figure 2:Image Analysis Workflow

C. Video Analysis Module

The video analysis module is designed to detect deepfake videos by analyzing both spatial and temporal information present in video data. Since a video consists of a sequence of frames, the system first extracts key frames at regular intervals to reduce computational complexity while preserving important visual information.Each extracted frame is processed using a Convolutional Neural Network (CNN) to identify visual inconsistencies such as unnatural facial expressions, irregular movements, and mismatched lip synchronization. These anomalies are common indicators of manipulated or synthetic video content.

In addition to frame-level analysis, the system considers temporal consistency across consecutive frames to improve detection accuracy. The predictions obtained from individual frames are aggregated to produce a final classification result for the entire video. This approach ensures a balanced and reliable detection process.



TruthCheck – Deep Fake and Fake News Detector

Video Deepfake Detection

Upload a video to detect face manipulation or AI-generated content.

Selected file: 0000.mp4

Prediction: REAL VIDEO

Explainable Results

Frames analysed : **20**

Fake frames : **0**

Real frames : **20**

Final Prediction : REAL VIDEO

The uploaded video was processed by extracting frames at regular intervals. A total of 20 frames were analyzed using the deep learning model. Out of these frames, 0 were classified as fake and 20 were classified as real. The final prediction is determined using majority voting across all analyzed frames. If most frames show manipulation patterns such as facial inconsistencies, unnatural blending, or temporal artifacts, the video is classified as a FAKE VIDEO. Otherwise, it is considered a REAL VIDEO.

Inference Time: 3.61s s

100% Truth Score

0% Fake Score

100% Model Confidence

Analyzed Frames

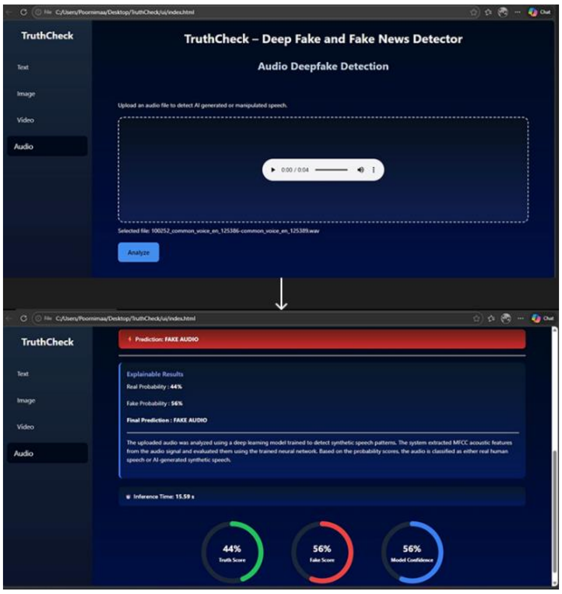
Figure 3: Video Analysis Workflow

D. Audio Analysis Module

The audio analysis module focuses on detecting synthetic or manipulated audio signals by analyzing their frequency and temporal characteristics. The input audio is first converted into a spectrogram representation, which provides a visual representation of frequency variations over time.

A deep learning model is used to extract features from the spectrogram and identify patterns associated with real and fake audio. The model analyzes characteristics such as pitch variation, frequency distribution, and signal consistency to detect irregularities in the audio signal.

Based on the extracted features, the system classifies the audio as real or fake and provides a confidence score. This module enhances the system's ability to detect deepfake speech and improves overall detection capability.



TruthCheck – Deep Fake and Fake News Detector

Audio Deepfake Detection

Upload an audio file to detect AI generated or manipulated speech.

Selected file: 100252_common_voice_en_125386-common_voice_en_125389.wav

Prediction: FAKE AUDIO

Explainable Results

Real Probability : 44%

Fake Probability : 56%

Final Prediction : FAKE AUDIO

The uploaded audio was analyzed using a deep learning model trained to detect synthetic speech patterns. The system extracted MFCC acoustic features from the audio signal and evaluated them using the trained neural network. Based on the probability scores, the audio is classified as either real human speech or AI-generated synthetic speech.

Inference Time: 15.59 s

Figure 4: Audio Analysis Workflow

E.Output Generation

The final stage of the system involves generating the output based on the predictions obtained from different modules. Each module processes its respective input and produces classification results indicating whether the content is real or fake.

The system presents the results in a structured and user-friendly format to improve interpretability. The output includes a truth score, fake score, and an overall confidence level, which helps users understand the reliability of the prediction.In addition to classification, the system ensures that the outputs are clear and easily understandable, allowing users to make informed decisions about the authenticity of the given content. This final stage plays an important role in enhancing transparency and usability of the TruthCheck system.

III. SYSTEM ARCHITECTURE

The TruthCheck system follows a well-structured architecture designed to support efficient and accurate multi-modal fake media detection. The system integrates multiple components including the user interface, processing units, and machine learning models, which work together to analyze and classify different types of digital content.
The architecture is designed to handle multiple input formats such as text, images, videos, and audio within a unified framework. Each type of input is processed through dedicated modules, ensuring that appropriate techniques are applied for analysis. This modular design improves scalability, flexibility, and maintainability of the system.
The overall workflow of the system ensures smooth data flow from input collection to final output generation. Each layer in the architecture performs a specific function, and proper communication between layers ensures efficient processing and accurate results.

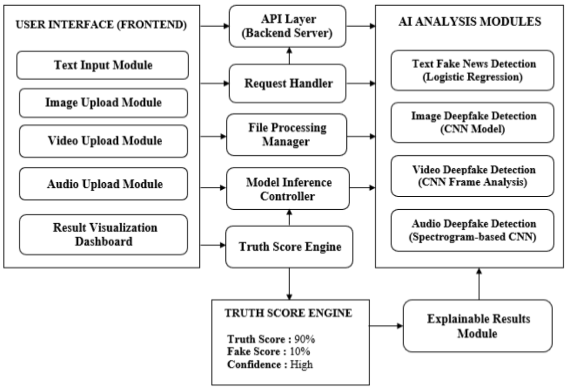

USER INTERFACE (FRONTEND)
Text Input Module
Image Upload Module
Video Upload Module
Audio Upload Module
Result Visualization Dashboard
API Layer (Backend Server)
Request Handler
File Processing Manager
Model Inference Controller
Truth Score Engine
AI ANALYSIS MODULES
Text Fake News Detection (Logistic Regression)
Image Deepfake Detection (CNN Model)
Video Deepfake Detection (CNN Frame Analysis)
Audio Deepfake Detection (Spectrogram-based CNN)
TRUTH SCORE ENGINE
Truth Score : 90%
Fake Score : 10%
Confidence : High
Explainable Results Module

Figure 8: Overall System Architecture of TruthCheck

A. Input Layer

The input layer acts as the entry point of the system, where users provide data in different formats such as text, images, videos, or audio files. The system interface is designed to accept multiple types of inputs, making it user-friendly and flexible for real-world usage.
This layer is responsible for validating the input data and ensuring that it is in the correct format before passing it to the next stage. It also handles file uploads and text inputs efficiently, allowing seamless interaction between the user and the system. By supporting multiple media formats, the input layer plays a key role in enabling multi-modal analysis within a single platform.

B. Processing Layer

The processing layer is responsible for preparing the input data for analysis. This includes preprocessing and feature extraction, which are essential steps in improving model performance.

Different preprocessing techniques are applied based on the type of input data. For text, cleaning operations such as removal of stop words and noise are performed. For images, resizing and normalization are applied. In the case of videos, frames are extracted at regular intervals, while audio data is converted into spectrogram representations.
After preprocessing, feature extraction techniques are used to convert raw data into meaningful representations. These features capture important patterns and characteristics required for accurate classification.This layer ensures that all input data is transformed into a consistent and suitable format before being passed to the model layer.

C. Model Layer

The model layer consists of machine learning and deep learning models used for detecting fake content. Different models are assigned for different types of media to improve accuracy.

- Logistic Regression is used for text analysis
- Convolutional Neural Networks (CNN) are used for image and video analysis

- Spectrogram-based deep learning models are used for audio analysis

Each model processes the extracted features and generates prediction results. This layer plays a crucial role in identifying patterns and detecting manipulated content.

D. Output Layer

The output layer is responsible for presenting the final results to the user. The predictions generated by the model layer are converted into a structured format that includes truth score, fake score, and confidence level.The results are displayed in a clear and understandable manner, allowing users to interpret the authenticity of the content easily. This layer ensures transparency and improves user trust in the system.

IV. RESULT AND DISCUSSION

The proposed system, TruthCheck, was successfully developed and tested for detecting fake media across multiple formats including text, images, videos, and audio. The system was evaluated based on functionality, performance, usability, and reliability under different test conditions. The results demonstrate that the system is capable of accurately identifying manipulated content while maintaining efficient performance.

A.Functional Testing Results

All modules of the system were tested individually and collectively to ensure proper functionality. The text analysis module was able to classify fake and real news content accurately based on trained datasets. The image analysis module successfully detected manipulated images by identifying visual inconsistencies such as artifacts and unnatural patterns.

The video analysis module effectively analyzed extracted frames and detected deepfake content by identifying temporal and spatial inconsistencies. The audio analysis module was able to detect synthetic speech by analyzing spectrogram features and identifying irregular frequency patterns.

The system generated prediction outputs including truth score, fake score, and confidence level for all types of input. These results confirm that the system functions correctly across different media types.

B.Performance Evaluation

The system demonstrated efficient performance during testing. Since lightweight machine learning and deep learning models were used, the system was able to run smoothly on CPU-based environments without requiring high-end hardware.The response time for processing different types of input remained within acceptable limits. Text and image processing produced faster results, while video and audio processing required slightly more time due to additional computations such as frame extraction and spectrogram generation.Overall, the system maintained stable performance even when handling multiple inputs, indicating good scalability and efficiency.

C.Usability Analysis

The system interface was designed to be simple and user-friendly, allowing users to easily upload or input different types of media. The output results were presented in a clear and understandable format, including prediction labels and confidence scores.

Users were able to interpret the results without requiring technical knowledge, which improves the practical usability of the system. The multi-modal capability also enhances user experience by allowing analysis of different media types within a single platform.

D.System Reliability and Security

The system showed reliable performance across all modules. The models were able to consistently detect fake content based on learned patterns and features. The use of multiple modules improved the overall detection capability of the system.The inclusion of confidence scores and structured outputs improved transparency and helped users understand the reliability of the predictions. The system also maintained consistency in results across repeated tests, indicating stability and robustness.

V. PERFORMANCE ENHANCEMENT

The performance of the TruthCheck system is improved through the use of efficient machine learning and deep learning models combined with optimized data processing techniques. Lightweight models such as Logistic Regression for text analysis and Convolutional Neural Networks (CNN) for image and video processing ensure that the system operates effectively even on CPU-based environments without requiring high-end hardware. This makes the system more accessible and practical for real-world applications.

The system also benefits from efficient preprocessing methods applied to different types of media. Techniques such as text cleaning, image normalization, frame extraction for videos, and spectrogram generation for audio help in improving the quality of input data. These steps enhance feature extraction and contribute to better accuracy in detecting fake content across multiple media formats.

Furthermore, the use of separate modules for handling text, image, video, and audio data improves processing efficiency and scalability. This modular approach reduces computation time and allows the system to handle multiple inputs effectively. As a result, the system achieves faster response time, improved accuracy, and reliable performance compared to traditional single-modal detection systems.

VII. CONCLUSION

The TruthCheck system provides an effective solution for detecting fake media across text, images, videos, and audio. By integrating machine learning and deep learning models, the system is able to analyze multiple types of content and generate accurate prediction results.

The use of lightweight models and efficient preprocessing techniques ensures smooth performance on standard systems. The inclusion of truth score, fake score, and confidence level improves transparency and helps users understand the results easily.

Overall, the system offers a reliable and scalable approach to reduce the spread of misinformation and supports real-world applications.

REFERENCES

[1] J. Pu, Z. Sarwar, S. M. Abdullah, A. Rehman, Y. Kim, P. Bhattacharya, M. Javed, and B. Viswanath, “Deepfake Text Detection: Limitations and Opportunities,” in Proc. IEEE Symp. Security and Privacy (SP), 2023, pp. 1–15.

[2] S. Alharbi, M. Khan, and A. Alotaibi, “Deepfake Detection: A Multi-Algorithmic and Multi-Modal Approach for Robust Detection and Analysis,” in Proc. IEEE International Conference, 2023.

[3] S. Khan and D. Dang-Nguyen, “Deepfake Video Detection: A Comprehensive Survey of Advanced Machine Learning and Deep Learning Techniques to Combat Synthetic Video Manipulation,” in Proc. IEEE Conference, 2024.

[4] S. Chakraborty and R. Ghosh, “FaND-X: Fake News Detection Using Transformer-Based Multilingual Masked Language Model,” in Proc. IEEE Conference, 2023.

[5] H. Singh and A. Mishra, “Deepfake-Speech Detection with Pathological Features and Multilayer Perceptron Neural Network,” in Proc. IEEE Conference, 2023.

[6] A. Roy, P. Das, and S. Dutta, “ETMA: Efficient Transformer-Based Multilevel Attention Framework for Multimodal Fake News Detection,” IEEE Access, 2023.

[7] M. Singh, R. Patel, and A. Verma, “Deepfake Detection with Deep Learning: Convolutional Neural Networks Versus Transformers,” in Proc. IEEE Conference, 2023.

[8] S. Islam, A. Rahman, and M. Hossain, “Bangla Fake News Detection Using Machine Learning, Deep Learning and Transformer Models,” in Proc. IEEE Conference, 2022.

[9] K. Barman and S. Dutta, “A Novel Fake News Detection Model for Context of Mixed Languages Through Multiscale Transformer,” IEEE Access, 2023.

[10] K. Verma and S. Gupta, “A Federated Convolution Transformer for Fake News Detection,” IEEE Access, 2023.

[11] M. Ali, R. Ahmed, and S. Rahman, “Fake News Detection Using Deep Learning and Transformer-Based Model,” in Proc. IEEE Conference, 2023.

[12] J. Li, H. Wang, and Y. Zhao, “Deepfake Detection Based on Multi-Scale RGB-Frequency Feature Fusion,” in Proc. IEEE Conference, 2024.

[13] A. Kumar, S. Gupta, and P. Sharma, “READFake: Reflection and Environment-Aware DeepFake Detection,” in Proc. IEEE Conference, 2025.

[14] R. Singh, P. Jain, and A. Patel, “DeepDect: A Facial Deepfake Video Detection Application Using Ensemble Learning,” in Proc. IEEE Conference, 2025.

[15] [M. Khan, A. Hussain, and T. Ali, “The Deepfake Dilemma: Enhancing Deepfake Detection with Vision Transformers,” *in Proc. IEEE Conference, 2025.*