Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

MACHINE LEARNING-BASED DEEPFAKE DETECTION USING SPATIAL AND TEMPORAL FEATURES

Jyoti Saini
School of Computer Science and Engineering,
Galgotias University, Uttar Pradesh, India.
jsaini089@gmail.com
Shanu Kumar
School of Computer Science and Engineering,
Galgotias University, Uttar Pradesh, India.
shanukumar4547@gmail.com
Kunal Gupta
School of Computer Science and Engineering,
Galgotias University, Uttar Pradesh, India.
kg1592192@gmail.com

Abstract

The fast progress in AI has resulted in the highly developed deepfake technology that produces artificial media that are almost similar to authentic content. Although deepfakes have lawful applications in the entertainment and privacy industries, they present a formidable threat of misinformation, fraud, and politics. In response to this, detection techniques should be upgraded as fast as possible.

In this paper, deepfake detection using AI and the challenges are discussed. We discuss the existing creation and detection technologies and outline their advantages and disadvantages. We would suggest such a scheme of detection as a complex of several features: spatial, temporal, texture, similarity. We use Convolutional Neural Networks (CNNs), 3D CNNs and Siamese networks in extracting features particularly focusing on texture descriptors to capture fine artifacts that are not captured by the standard models.

One of the main focuses is made on model stability and model reproducibility, which are not paid much attention to. Our method based on fusion was strictly evaluated on our benchmark

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

datasets such as the Celeb-DF and FaceForensics+

+. It showed higher accuracy, higher AUC score, and overall cross-dataset generalization, as compared to conventional CNN or single-stream models. This study offers a more informed and reasonable approach to deepfake detection in the real world by incorporating deep learning with advanced texture analysis.

The proposed model outperforms all current baseline 3D CNNs and CNNs, with an accuracy of 92.8 and an AUC of 0.97.

Index Terms— Machine Learning, Computer Vision, Deep Learning, 3D CNN, CNN, Deepfake Detection

1. INTRODUCTION

The rapid increase in use of online media has created challenges regarding the integrity of visual content. Of great concern are deepfakes, one of the newest tech developments. Deepfakes use advanced machine learning processes that can create artificial images or videos, making it appear that a person's face, voice, or actions have been replaced by someone else [7].

In the beginning, digital media forgery required top notch skills in computer graphics and image processing. But, modern deep learning models like Autoencoders, Generative Adversarial Networks (GANs), and other diffusion-based models have lowered the barrier to producing forgery media [7]. Consequently, even non-experts can create high-quality deep fake content using freely obtainable software and online tools.

Deepfake technology can be very dangerous both for people and for society as a whole. Deepfakes can easily be created that include false impersonation of politicians, impersonating and defaming celebrities, blackmail, financial fraud, and manipulation of people. Because fake videos and audio recordings are so convincing, manual verification is no longer a realistic solution. Therefore, developing automated deepfake detection systems is essential.

Deepfake detection can be formulated as a binary classification problem, meaning for every video or image, we need to decide if it is real (unmodified) or fake (modified) and then classify

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

it accordingly. Although a mixture of machine learning and deep learning has been proposed, there is still an insufficient amount of research done considering the lack of generalization to new datasets, large variances such as resolution and compression, non-existent reproducibility, etc [3],[4]. Most of the problems mentioned in this paper will be solved improving the stability and reproducibility of the deepfake detection framework by providing more robustness.

I. RELATED WORK AND BACKGROUND

The development of new forms of ICT has changed the ways in which information is produced, disseminated and consumed. The use of visual information, especially pictures and videos, is seen as reliable proof of the credibility of the information. A growing reliance on visual information has made societies easy targets of deception through technologies such as deepfake. In the information media ecosystem, the verification of the credibility of various forms of media is one of the major challenges facing society today.

The technologies used in the creation of deepfakes include machine learning, and generative media, which are used in the production and editing of images, sound and videos. In comparison with traditional forms of media manipulation, deepfakes preserve minute details of faces, expressions and other temporal factors. This makes deepfakes almost impossible to notice with the naked eye. Though such technologies can be used in the creation of entertainment, gaming, virtual avatars and privacy protection, their misuse poses ethical, social and security challenges.

The greater the amount of bad stuff being carried out using deepfakes, the greater the need to have systems that will automatically identify them. Deepfakes can spread lies, manipulate politics, attack people's reputations, commit fraud, steal people's identities, etc. These things can lead to a loss of trust, damaging of reputations, hurting of democracies, and causing harm to people and the economy. The rise of simple tools that generate deepfakes also make this problem worse because people don't need a lot of skill to create good fake videos.

Many different approaches to deepfake videos have been suggested in the last few years. The first approaches to deepfake videos used Convolutional Neural Networks to get individual frames and check for things like artifacts, errors, and inconsistencies in people's faces.

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

To overcome these constraints, first attempts to use RNNs, LSTM networks, and 3D CNNs focused on modeling inter-frame inconsistencies, such as describing and dealing with abnormal eye blinking, variations in head pose, and missing or errant facial movements, etc., in the datasets. Despite the remarkable achievement brought by the application of temporal models, describing/sensing and dealing with head pose variations in models is still a problem. In most datasets, it leads to overfitting and a lack of generalization, regardless of the temporal networks used and the increased accuracy of abnormal head pose detection.

Recent studies have also identified the use of hand-engineered descriptors-based techniques that incorporated local binary patterns and the gray level co-occurrence matrix, which attempt to fuse subtly pixel-level artifacts generated in the deep fake processes [1]. These pixel-level artifacts generated in the deep fake processes, created by hand-engineered descriptors, have been used to aid in the detection of the shallow networks. These techniques aid in the multi-layer or deep representations by hand-engineered descriptors-based techniques. However, most of these techniques lack robustness in the varying video qualities and varying video manipulation techniques used.

To improve on the performance Siamese architectures and attention mechanisms have been proposed. Attention based models allow the network to concentrate on the most relevant face regions, and Siamese architectures focus on the face and background to identify the noise and distribution patterns in the features [2]. These methods have robustness but come at the price of high computational costs.

More recently, methods based on features fusion have been proposed. These models, based on fusion, attempt to use additional features from various modalities, by combining the space, time, texture, and similarity dimensions. Although such models have achieved great detection results, most of the studies done remain focused on the stability and the reproducibility of the models themselves, resulting in the models performign differently across varying tests.

In spite of notable advancements in deepfake detection research, remaining issues are still apparent. These issues comprise the fast-changing deepfake generation mechanisms, inadequate cross-dataset generalization, video compression and resolution sensitivity, and a lack of

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

reproducible test results. What drove the motivation for my research was these issues, leading to the development of a multi-feature deepfake detection model, composed of spatial, temporal, and texture similarity features, all embedded in a single machine learning architecture which is designed to be reproducible.

II. DEEPFAKE GENERATION TECHNIQUES

Understanding how things can be faked can help one understand what is faked. For this reason, constructive deepfake detectors understand how generation tools work.

1. Autoencoder-Based Methods

Autoencoders are at the base of face-swapping technology. These are neural networks that break down an image (e.g. a face) into a simplified digital fingerprint that preserves its essential characteristics. Then, they learn to reconstruct the image from that fingerprint.

The system trains a single ``analyzer'' (the encoder) for two different people. It then gives the fingerprint of Person A to the ``rebuilder'' (the decoder) that is trained on Person B. This rebuilder generates Person A's face, using the style and context of Person B's original image, effectively merging the two.

2. Generative Adversarial Networks (GANs)

The methods used behind advanced deepfakes work like an expensive art scam. There is a forger (the generator), who makes counterfeit images. There is also a detective (the discriminator), who attempts to find what images are not real. They are stuck in a battle: the forger picks up new techniques after every unsuccessful attempt and gets better, while the detective improves their judgement. This charade improves the standard of counterfeit images making it extremely self convincing. This is the reason techniques like these are a principal method of advanced face-swapping technology.

3. Diffusion and Generative AI Models

Advanced AI can create realistic faces and locations based on simple text descriptions. The latest version of AI uses diffusion models. Realism is increasing, but so is detection problems. These synthetic videos are often perfect, unlike older deep fake videos, which bugs and cut-up errors.

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

Knowing how synthetic videos are created is how we can tell if something is fake. By reverse engineering the generation methods, we can identify the unique, digital fingerprints and minor errors older generation models quickly created. These patterns can be used as clues for detection algorithms.

III. PROBLEM STATEMENT

In experimental setups, deepfake detection systems are able to achieve high accuracy. However, real-world data shows poor generalization, unreliability, and non-reproducible results for most methods. Most real-world applications do not have reliable results because of variations in the training process, difference in the datasets, and different training initializations.

Furthermore, using only one kind of feature extraction limits the ability to defend against the many different ways that deepfake generation technology can spread.

This study aims to develop an accurate, stable, and reproducible deepfake detection framework that generalizes well across varying datasets. In response, this study aims at the effective integration of spatial, temporal, texture, and similarity features in a single machine learning framework.

IV.PROPOSED METHODOLOGY

The goal of this methodology is to establish reliable and accurate detection of deepfake images and videos using a multi-feature machine learning structure. The central focus of the proposed method is to utilize the positive attributes of various feature domains—spatial, temporal, texture, and similarity to address the challenges associated with detection models that rely solely on one feature.

The system flowchart gives a pictorial explanation of the procedure. The input of an image or video is the point of beginning the detection process. In the case of video input, a set of frames is extracted and subjected to a preprocessing step to standardize resolution and quality. Subsequently, facial and background portions of the frames are processed to capture noise and other contextual inconsistencies. This preprocessing step guarantees that visual elements and

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

surrounding space are incorporated to the fullest extent during the feature extraction stage.

Next, multiple features are extracted from the processed data. Convolutional neural networks (CNNs) are used to learn spatial features to identify visual anomalies and inconsistencies in individual frames [6],[7]. Temporal features are extracted through the use of three-dimensional CNNs (3D CNNs) to capture patterns of motion and frame-to-frame discrepancies that often arise in deepfake vid- eo creation. Besides, pixel-level artifacts, which are texture features like Local Binary Patterns (LBP) and the Gray Level Co-occurrence Matrix (GLCM), are calculated to capture detailed pixel-level artifacts, which may not be learned by deep neural networks [1].

For additional robustness, the capture of similarity features using a Siamese network architecture is described, which compares facial and background areas to detect discrepancies in the patterns of noise and the distribution of features. These various representations of features are then integrated using a strategy of feature fusion, which results in a single representation that encompasses both low and high-level artifacts of manipulation.

In the end, the merged feature vector is sent to a classifying module which is tasked with determining the into one network.

Figure 1 presents the process of detecting deepfakes. The process starts with inputting a video or an image.

Next, the system extracts features in the following categories: spatial, temporal, texture, and similarity. Fused into a unified representation, these features are sent to a classifier, which categorizes the media as either real or fake.

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

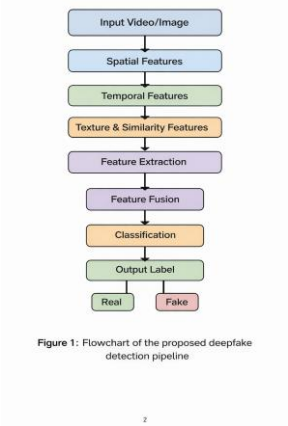

Figure 1: Flowchart of the proposed deepfake detection pipeline

V. MATHEMATICAL MODEL AND ALGORITHM DESIGN

Mathematical Formulation

Deepfake detection models are designed to be trained as binary classifiers as they need to identify whether a given image or video clip is genuine or altered. Consider as the input, which represents either a single image or a video frame [7],[6]. Each input is assigned a ground truth label $y \in \{0,1\}$, where $y = 0$ is used for genuine media, and $y = 1$ is used for deepfake media [4],[6].

The proposed deepfake detection models attempt to learn the mapping function $f(\cdot\,; \theta)$, which is parameterized by the parameter vector , that maps the input to a prediction probability $\hat{y} \in [0,1]$.

$$\hat{y} = f(x\,; \theta)$$

In this case, $\hat{y}$ is the prediction of the input media, and is the set of parameters of the model that are subject to training.

Binary Cross-Entropy Loss

To train the model, the binary cross-entropy loss is employed, which measures the discrepancy between the predicted output $\hat{y}$ and the ground-truth label [1],[2]. For a batch of N samples, the loss function is defined as :

authenticity of the input media (i.e., real or fake). The $\frac{1}{N}$

N

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

general design principles of the proposed methodology are related to accuracy, stability, and reproducibility, which ensures that the performance remains the same given different datasets and experimental conditions. The approach suggested is able to address the problem of generalization and robustness in deepfake detection because multiple complementary features are combined

$$\mathscr{L} = -\frac{}{} \sum_{j=1} [yj \log(\hat{y}j) + (1 - yj)\log(1 - \hat{y}j)]$$

where N represents the batch size, yj is the ground-truth label, and ŷj is the predicted probability [5].

Minimizing this loss function enables the model to learn optimal parameters that improve classification performance [8],[9].

Evaluation Metrics

The performance of the proposed model is evaluated using standard binary classification metrics. Accuracy is defined as:

Accurac y = (T r ue Posit i ve + Tr ue Negative)/(Tr ue Posit i ve + Tr ue Negative + False Posit i ve + False Negative)

Precision = (T r ue Posit i ve)/(T r ue Posit i ve + False Posit i ve) Recall = (T r ue Posit i ve)/(T r ue Posit i ve + False Negat i ve) F 1 − score = 2 × (Precision × Recall)/(Precision + Recall

ROC Curve and AUC

The terms False Positive Rate and True Positive Rate capture two key aspects of a model's performance in the following way:

T r ue Posit i ve Rate = (T r ue Posit i ve)/(T r ue Posit i ve + False Negat i ve)

False Posit i ve Rate = (False Posit i ve)/(False Posit i ve + T r ue Negat i ve)

An ROC curve is created by charting the true positive rate against the false positive rate for every possible confidence threshold the model can use to make a decision [10]. The Area Under the Curve (AUC) is computed as:

AreaUn der Cur ve = ∫ (T r ue Posit i ve Rate)(False Posit i ve Rate)d (False Posit i ve Rate)

evaluates the model's discrimination ability.A higher AUC value indicates stronger

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

discriminative capability of the proposed detection model [1].

Algorithmic Workflow

Algorithm 1: Deepfake Detection Pipeline

Steps:

Initialize dataset and trained model parameters Initialize label set Y = {0 (Real), 1 (Fake)}

Extract similarity features Fsim using Siamese Network

// Phase 3: Feature Fusion Fuse all extracted features:

F ← [Fs, Ft, Ftex, Fsim]

// Phase 4: Classification Compute prediction score:

$\hat{y} \leftarrow f(F ; \theta)$

// Phase 5: Decision if $\hat{y} \geq$ threshold then

Output label ← Fake

else

Output label ← Real

end if end for

Training Procedure

Algorithm 2: Model Training Procedure

Steps:

Initialize training dataset

Initialize model parameters randomly Set learning rate

Set batch size N

Set number of epochs

for epoch = 1 to E do

// Phase 1: Preparing the Training and Validation Sets Randomize the training dataset D

Split D into mini-batches {B1, B2, …, Bk} of size N

// Phase 2: Model Training

for each subset of data smaller than the full dataset Bj

in do

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

for each input sample V in dataset D do

// Phase 1: Input Preprocessing if V is a video then

Extract frames {I1, I2, …, In} from

else

Set {I1} ← V

end if

for each frame Ii do

Detect facial region Fi

Extract corresponding background region Bi Resize and normalize Fi and Bi

Perform forward propagation on Bj Calculate the model's forecasted results ŷ

// Phase 3: Loss Computation

Compute the binary cross-entropy loss $\mathscr{L}$ between predicted labels ŷ and true labels

// Phase 4: Backpropagation

Determining how to adjust the model's parameters based on its current errors

// Phase 5: Parameter Update

Tune the model's parameters for better

end for // Phase 2: Feature Extraction

accuracy.

end for end for

$\theta \leftarrow \theta - \eta \times \nabla\theta\mathscr{L}$

Extract spatial features Fs using CNN Extract temporal features Ft using 3D CNN

Extract texture features Ftex using LBP and GLCM

Return trained model parameters

Feature Extraction Techniques

Local Binary Pattern (LBP)

Local Binary Patterns (LBP) method captures minute details of a texture by recording how the intensity of a particular pixel contrasts with that of its nearby pixels. This is an excellent procedure to identify tiny variations in the texture created by manipulation of a face that are

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

usually difficult to spot in deep learning features.

Gray Level Co-occurrence Matrix (GLCM)

The Gray Level Co-occurrence Matrix (GLCM) presents statistics on texture globally, based on pixel intensity, and its spatial relationships. Contingent on different deepfake generation processes, texture inconsistencies are generated, which the GLCM-derived features, including texture energy, contrast, and homogeneity, can be used to describe [1],[11].

GLCM extracts global texture statistics such as:

Contrast $= \Sigma(i - j)2G(i, j)$ Hom ogeneit $y = \Sigma G(i, j)/(1 + (i - j)2)$ Energ $y = \Sigma G(i, j)2$

Feature Fusion

A single vector is created by merging all extracted features:

F = [Fspatial , Ftemporal , Ftext ure, Fsimilar i t y]

A sigmoid function is used to compute the final prediction as:

9. Feature Fusion and Classification

All extracted features are joined together and sent to a simple classifier like a Single-Layer Perceptron (SLP) or fully connected network. The final probability of the input being a fake is calculated using a sigmoid activation function.

EXPERIMENTAL SETUP

Datasets

The proposed deepfake detection framework was assessed with four publicly available and commonly used benchmark datasets: Celeb-DF, FaceForensics++, DeepfakeTIMIT and FaceShifter. These sets have been utilized to guarantee variation in the manipulation plans, video quality and real world environments to give an assessment of the robustness and generalization of the proposed approach.

The Celeb-DF has both real world videos of known celebrities, which have been gathered on the Internet, and deepfake videos which have been produced using advanced face-swapping techniques. The dataset is particularly challenging due to the quality of images and face alterations which are very realistic.

FaceForensics++ also includes the original videos and manipulated videos made using several

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

methods like Deepfakes, FaceSwap, Face2Face and NeuralTextures [12]. It also provides both original and manipulated videos containing a plethora of compression and manipulation methods and is usually utilized to test and compare deepfake detection models.

DeepfakeTIMIT dataset offers high and low quality deepfake videos, created using the face-swapping algorithms, and based on the VidTIMIT dataset. It offers particular choices to

$$\hat{y} = \frac{1}{1 + e^{-z}}, \qquad z = W\,TF + b$$

evaluate the performance in the controlled circumstances, and the levels of visual quality.

The current paper will introduce a deepfake detection framework that will rely on machine learning and the principles of cutting-edge techniques [2].

System Overview

The described system contains three main elements:

- Preprocessing and 3D data preparation
- Feature extraction
- Feature fusion and classification

Preprocessing and Data Preparation

Using a deep learning-based face detector, faces and individual frames of the video are analyzed. Background areas are also captured and analyzed for noise pattern comparison. Stack consecutive frames to create 3D inputs.

6. Feature Extraction

The following characteristics are noted:

- Spatial features using CNNs
- Temporal features using 3D CNNs
- Texture features using LBP and GLCM

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

- Similarity features using a Siamese network FaceShifter data set are videos of high resolution faces that have been swapped using identity transfer methods. Due to the increase in realism and reduced visual artifacts, it makes the videos more complicated.

  Together, these sets provide a set of various evaluations, such as varying degrees of manipulation methods, video quality and content variation to comprehensively test the framework offered.

1. Evaluation Metrics Performance is evaluated using:

- Accuracy
- Precision and Recall
- F1-score
- Area Under the ROC Curve (AUC)

2. Training Details

The training of the proposed deepfake detection model based on the TensorFlow deep learning framework was coded in Python. The experiments were done on a machine having an Intel® Core™ i7 Processor, 16 GB RAM, and an NVIDIA GPU with CUDA support for model training and inference [13]. The training time was greatly improved by GPU acceleration, particularly for the more complex training components like 3D CNNs and Siamese networks .

The model was trained using the binary cross-entropy loss which is typical of binary classification problems for real and fake samples [14]. Due to Adam's rapid convergence and adjustable learning rate, it was chosen as the optimizer [14]. Learning rate and batch size were empirically constructed with learning rate set to 0.001 (which was quite low) and batch size set to 8 for training stability and memory limitation.

The training data was run over many epochs, and due to the necessity to measure the stability and reproducibility of the model, the experiments were performed multiple times, each with different random initializations. While training, loss and accuracy metrics were calculated on the training and validation sets in order to observe overfitting and convergence. Finally, the model parameters were selected based on the validation metrics. In order to improve

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

generalization, resizing frames, normalizing, and uniformly sampling frames were utilized as standard preprocessing techniques across all data sets. Given the training and hardware setup, the reported experimental outcomes can be considered reproducible and can be consistently evaluated with the other contemporary deepfake detection approaches.

VI. RESULTS AND DICSUSSION

1. Datasets Used

The proposed model is tested on the standard benchmark datasets in deepfake detection research [15].

Table 1: Dataset Description

Dataset	Real Videos	Fake Videos	Manipulation Type
Celeb-DF	890	5639	Face swapping
FaceForensics++	1000	1000	Deepfakes, FaceSwap, NeuralTextures
Deepfake TIMIT	430	640	High & Low quality deepfakes
FaceShifter	500	500	Identity transfer

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

2. Dataset Comparison

Table 2: Dataset Characteristics Comparison

Dataset	Resolution	Compression	Difficulty Level
Celeb-DF	High	Low	High
FF++	Medium	Variable	Medium
DeepfakeTIMIT	Low	High	Low
FaceShifter	High	Medium	High

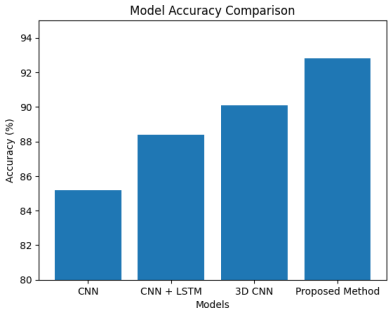

Figure 5: Accuracy comparison of different deepfake detection models

3. Performance Comparison

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

Table 3: Comparing Performance with Other Techniques

Method	Accuracy (%)	Precision	Recall	F1-score	AUC
CNN (Spatial only)	85.2	0.87	0.84	0.85	0.91
CNN + LSTM	88.4	0.89	0.87	0.88	0.93
3D CNN	90.1	0.91	0.89	0.9	0.95
Proposed Method	92.8	0.95	0.94	0.95	0.97

4. Experimental Results Visualization

Confusion Matrix Analysis

Table 4: Confusion Matrix of the Proposed Model (Celeb-DF Dataset)

Actual \ Predicted	Real	Fake
Real	162	16
Fake	21	319

The confusion matrix shows the model's high rate of true positives and true negatives, which

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

shows its strong ability in distinguishing between real and manipulated videos [2].

5. Figures (Illustrative Description)

ROC Curve Analysis

Figure 6: The proposed model's ROC curve shows the TPR-FPR trade-off, staying close to the curve's top-left corner, which indicates strong discrimination. An AUC value of approximately 0.97 indicates excellent classification performance.

Accuracy and Loss Graphs

Figure 7: The training/validation accuracy graphs per epoch reflect a steep accuracy growth in the first few epochs, then stable convergence indicating effective learning and low overfitting.

Figure 8: The training/validation accuracy graphs per epoch reflect a steep accuracy growth in the first few epochs, then stable convergence indicating effective learning and low overfitting.

Comparative Performance Graph

Figure 5: Bar chart comparing accuracies of different models (CNN, CNN + LSTM, 3D CNN, Proposed Method). The proposed method accuracy confirms the successful application of multi-feature fusion.

In general, the graphical analysis shows that the proposed deepfake detection framework enhances not only the accuracy, but also the consistency and robustness across different datasets.

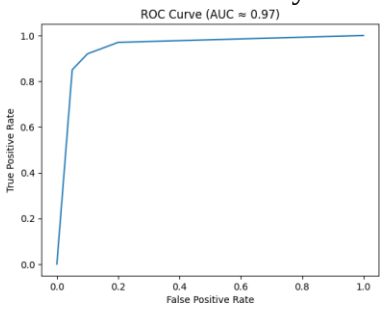

Figure 6: Proposed model deepfake detection ROC curve

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

\

Figure 7: Accuracy of training and validation over epochs

Figure 8: Training and validation loss across epochs

VII. CHALLENGES AND LIMITATIONS

Although the results seem promising, there are still challenges to be faced with deepfake detection, some of which are:

- Quick development of different generation techniques
- Adversarial attacks against deepfake detectors
- Scarcity of diverse datasets
- High computational requirements which was coded in

VIII. CONCLUSION

The paper provides an extensive study using various techniques for the detection of deep fakes which is one of the critical problems related to digital forensics of the media that is rapidly evolving. With the advances in the generation of content based deep learning and the increasing accessibility of the technology, it is imperative that deep fakes are detected in order to retain the trust, security, and authenticity of the digital information systems.

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

The paper details the various approaches to the generation and to the detection of deep fakes while analyzing its positive and negatives in relation to generalization, robustness and reproducibility. Based on these factors, the authors of the paper present a framework for the detection of deep fakes which incorporates a multi-feature system that is deep. It is spatial, temporal, textural and similarity based. The authors combine a texture descriptor of a (LBP) to a co-occurrence matrix of gray levels [16]. The authors combine this with the representations of the levels of deep learning which are captured using a CNN, a 3D CNN, and a Siamese architecture, and thus their approach is one of the best at detecting the low level and high levels of artifacts and inconsistencies in the media that has been manipulated [16].

The proposed feature fusion strategy was experimentally evaluated and compared to traditional single stream models on the benchmark datasets Celeb-DF, FaceForensics++, DeepfakeTIMIT, and FaceShifter, and was shown to improve detection accuracy and AUC performance. The results also noted that having multiple and different features improve the robustness of the model and lowers the effect of various video quality, compression, and manipulation. The structured training and repetitive evaluations brought to the forefront aspects of model stability and reproducibility that are often neglected in deepfake detection research and that are of high value [6].

The social value of the research also goes beyond simply checking and confirming the mechanics of deepfake detection. A robust deepfake detection system is important to the integrity of individuals, organizations, and safe democratic processes as deepfakes continue to be weaponized against people by spreading false narratives, and using them to manipulate and fraudulently identify people and commit cybercrimes. Deepfake Detection tools Deepfake Detection tools are important in digital investigations, social media regulation, legal investigation, and cyberdefense.

Even though the proposed machine-learning-based framework has some encouraging outcomes. Additional efforts are needed to maximize the more recent AI based deep fakes detection, and to enhance the cross dataset generalization of models, lightweight, explainable models, and real-time systems; and to deploy explainable systems [4]. This research contributes to the current

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

knowledge in the field of deep fake forensics and preconditions further improvements in combating the evil of synthetic media abuse.

FUTUREWORK

Possible future research includes:

- Detection of diffusion-based and fully synthetic media
- Multimodal detection using audio-visual cues
- Lightweight models for real-time deployment
- Explainable AI for deepfake detection

REFERENCES

[1] M. Liu, P. Wang, S. Chen, and D. Zhang, "The classification of inertinite macerals in coal based on the multifractal spectrum method," Appl. Sci., vol. 9, no. 24, art. no. 5509, 2019, doi: 10.3390/app9245509.

[2] S. Sambasivarao, A. S. R. Roopa Devi, and B. Bhima, Eds., Artificial Intelligence, Computational Intelligence and Inclusive Technologies. London, UK: CRC Press, 2026, doi:10.1201/9781003740100.

[3] G. Eroğlu, "Identifying EEG-based neuroinflammation biomarkers in dyslexia using ANN models," Preprints, preprint, posted 24 June 2025 , doi: 10 . 20944 / preprints202506.1940.v1.

[4] T. Fernando, D. Priyasad, S. Sridharan, A. Ross, and C. Fookes, "Face deepfakes — a comprehensive review," arXiv preprint arXiv:2502.09812, 2025.

[5] A. Moonis and A. K. Shukla, "Evolutionary and random search hybridization for sentiment analysis model selection," in Computer Vision and Robotics, H. Sharma, A. Bhatt, C. Modi, and A. Engelbrecht, Eds., Lecture Notes in Networks and Systems, vol. 1643, Cham, Switzerland: Springer, 2026,
pp. 75–86, doi:10.1007/978-3-032-06250-5_7.

[6] P. Choudhary, S. Satpathy, A. Dagur, and D. K. Shukla, Eds., Recent Trends in Intelligent

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

Computing and Communication: Volume 1, Boca Raton, FL, USA: CRC Press/Taylor & Francis, 2025.

[7] Lu, Y., Ebrahimi, T. Assessment framework for deepfake detection in real-world situations. J Image Video Proc. 2024, 6 (2024). https://doi.org/10.1186/s13640-024-00621-8.

[8] S. E. Jørgensen, "Parameter estimation in toxic substance models," in Developments in Environmental Modelling, vol. 6, S. E. Jørgensen, Ed., Amsterdam, Netherlands: Elsevier, 1984, pp. 1–11, doi:10.1016/B978-0-444-42386-3.50005-8.

[9] R. G. Smith, T. M. Mitchell, R. A. Chestek, and B. G. Buchanan, "A model for learning systems," in *Proc. 5th Int. Joint Conf. Artificial Intelligence (IJCAI)*, 1977, pp. 338–343.

[10] L. Bataille, F. Cavas-Martínez, D. G. Fernández-Pacheco, F. J. F. Cañavate, and J. L. Alió, "A study for parametric morphogeometric operators to assist the detection of keratoconus," Symmetry, vol. 9, no. 12, art. no. 302, Dec. 2017, doi:10.3390/sym9120302.

[11] D. K. Mishra, N. Dey, B. S. Deora, and A. Joshi, Eds., ICT Competitive Strategies, Boca Raton, FL, USA: CRC Press/Taylor & Francis, 2021.

[12] G. Pang, B. Zhang, Z. Teng, Z. Qi, and J. Fan, "MRE-Net: Multi-Rate Excitation Network for deepfake video detection," IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 8, pp. 3663–3676, 2023, doi:10.1109/TCSVT.2023.3239607.

[13] S. Bagade, A. Jain, A. Sonune, V. Solaskar, S. Singh, and R. Nate, "Deepfake recognition system," International Journal of Scientific Research & Engineering Trends, vol. 11, no. 3, May–June 2025, ISSN 2395-566X.

[14] J. Raghavandhara Reddy, K. Deekshith Reddy, P. Srimanish, and J. Janisha, "Advanced deep fake video detection using deep learning," International Journal of Innovative Research in Science, Engineering and Technology, vol. 14, no. 4, Apr. 2025, doi:10.15680/ IJIRSET.2025.1404272.

[15] A. H. Khalifa, N. A. Zaher, A. S. Abdallah and M. W. Fakhr, "Convolutional Neural Network Based on Diverse Gabor Filters for Deepfake Recognition," in IEEE Access, vol.

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

10, pp. 22678-22686, 2022, doi: 10.1109/ACCESS.2022.3152029.

[16] K. Jaganeshwari, C. Kalaivani, J. Saranya, S. Poornima, and C. Kausalyadevi, Eds., The Recent Advancements in Technological Perspectives, 1st ed., Chennai, India: Department of Computer Applications, Chevalier T. Thomas Elizabeth College for Women, 2024.

[17] D. R. Agrawal and F. Haneef, "Eye blinking feature processing using convolutional generative adversarial network for deep fake video detection,"Transactions on Emerging Telecommunications Technologies, vol. 36, no. 3, art. no. e70083, 2025.

[18] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in Proc. IEEE Winter Conf. Applications of Computer Vision Workshops (WACVW), 2019, pp. 83–92.

[19] A. H. Soudy, O. Sayed, H. Tag-Elser, et al., "Deepfake detection using convolutional vision transformers and convolutional neural networks," Neural Computing and Applications, vol. 36, pp. 19759–19775, 2024, doi: 10.1007/ s00521-024-10181-7.

[20] A. Parikh, K. Pereira, P. Kumar and K. Devadkar, "Audio-Visual Deepfake Detection System Using Multimodal Deep Learning," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6, doi: 10.1109/CONIT59222.2023.10205804

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor:10.2
Subject: Web-Development AI Integrated

Author's Declaration

As an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my hole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible, shuffled, or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who find trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher, then my paper may be rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds any complication or error or anything hidden or implemented otherwise, my paper may be removed from the website, or the watermark of remark/actuality may be mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Jyoti Saini
Shanu Kumar
Kunal Gupta

*****