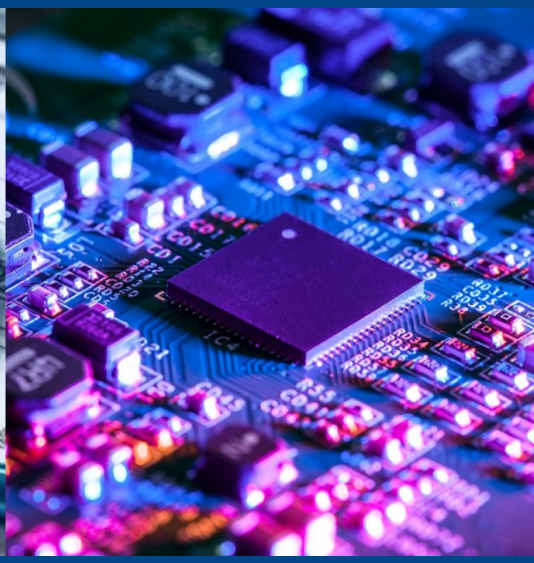
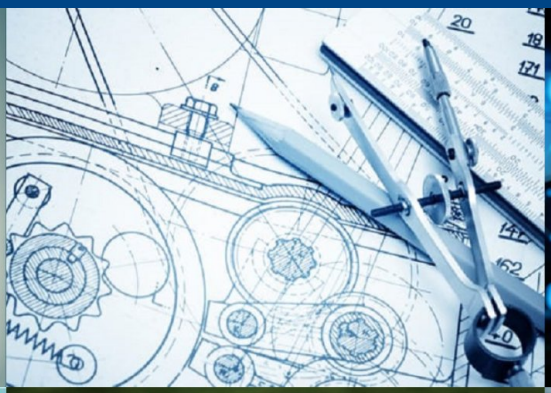
ISSN: 2582-7219

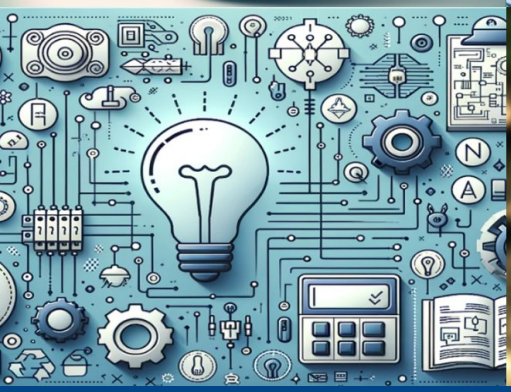




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026| DOI:10.15680/IJMRSET.2026.0904034

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deepfake Detection using Machine Learning

Keesara Sreya[1], Nagella Mythili[2], Kanneboina Renuka[3], Kanumuri Keerthi[4]

U.G Students, Department of Electronics and Communication Engineering, R.V.R&J.C College of Engineering, Chowdavaram, Guntur, India[1,2,3,4]

ABSTRACT: With the advancement of artificial intelligence, deepfake technology has rapidly progressed, allowing highly realistic manipulated videos and images to be created. Deepfakes are useful for entertainment and media but they can lead to severe issues like misinformation, identity theft, and digital fraud. The project is based of a Convolutional Neural Network (CNN) for a deepfake detection system that automatically classifies real vs fake videos. A video dynamic approach can involve the extracting of frames from videos, face detection and cropping, image preprocessing to create a structured dataset. You will then use this processed data to train a CNN model that learns spatial and texture -based features in order to detect manipulation artifacts. The performance of the model in detecting deepfake content is evaluated using accuracy precision, recall and F1-score metrics providing reliable detection of a given video as a deepfake or not. This system pilots enhancing the authenticity of digital media, a key step in combating misinformation.

KEYWORDS: Deepfake Detection, Convolutional Neural Network(CNN), Machine Learning, Deep Learning, Computer Vision, Face Extraction, Image Classification, Video Frame Analysis, Binary Classification , Digital Media Security.

I. INTRODUCTION

Deepfake technology is a powerful use of artificial intelligence that creates extremely realistic fake videos and images by altering faces[1]. As social media networks, and the communication possibilities of a digital world continue to explode, multimedia content is reaching new levels of shareability[2]. Deep fake technology can be used positively in the domains of entertainment, social media, movie-making, and virtual reality; however it is also being misused which has raised serious concerns about its usage related to misinformation or disinformation, identity theft, cyber-crime and digital fraud[3]. It is extremely challenging for humans to detect deepfakes manually[4] because modern generation techniques of the deepfake videos make them seem realistic and natural to human comprehension. Nonetheless, such forged videos typically have subtle artifacts and inconsistency that easily escape human perception[5]. Machine learning and deep learning methods are a promising way of tackling this problem because they can automatically learn patterns from the input data, thereby allowing efficient detection of subtle traces of manipulation within an image or video[6]. Some of these techniques include Convolutional neural network(CNNs) which are known to be one of the most successful algorithms for image and video analysis given their feature extraction capability[7].

This project is a CNN based deepfake detection system that will automatically classify all the videos are real or fake[8]. First, it extracts frames from videos, detects and crops the faces and preprocesses the image to form a processed dataset[9]. It then trains the CNN model to extract features from a given input image which can be either real0 or manipulated[9]. We evaluate the model using performance metrices like accuracy, precision, recall, and F1-score[11]. This project aims at said

II. LITERATURE SURVEY

The advent of image editing tools and application for public use increases the availability and accessibility of them, which causes a major challenge known as Fake image detection using Machine Learning. Several approaches utilizing machine learning techniques have been developed in recent years. This literature review reviews the current findings on this subject and points out difficulties as well as future pathways.

Farid and Lyu (2004) are among the first to conduct a machine learning-based study focusing on fake detection in images [13], they presented a statistical method to detect digital image forgeries through the analysis of inconsistencies in image properties. Since those days, a considerable number of machine learning based approaches have been

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

proposed to detect fake images including deep learning models such as convolutional neural network (CNNs), generative adversarial networks (GANs) and recurrent neural network (RNN).

A study by Nguyen et al. in 2017 Recently, Wang et al. (2017) [14] proposed a GAN detection method based on classical convolutional networks. The authors employed a variation of the GAN framework to create synthetic images, then employed a CNN to classify these generated images as true or false. The method proposed in that study outperformed other state-of-the-art models on this
In 2018, a study by Lietal. (2018), [15] to detect image splicing, which is combining two or more images together A CNN was used for characterizing the features of original and blended images and subsequently a decision tree was applied to classify the images into original or blended category.

Another study by Li et al. proposed a hybrid model, which incorporates the power of CNNs and RNNs for both image splicing detection/manipulation[16]. The proposed method first extracted image feature using CNN and then analyzed the temporal dependencies with RNN.

This approached was studied in 2020 by Huh et al. (2020),[18] used method for detecting deep fake videos trained with CNN/ GAN. This method involved a CNN being used to extract features from every frame of the video and then creating fake videos by using GANs. In this method, first a fake video is produced and then compared with the real to find any differences.

Although machine learning has made some inroads into detecting fake images from very few training samples, there are still many challenges and limitations. A key issue is the development of strong models that can realize advanced image modifications. Also to work properly, machine learning models need thousands and millions of real pictures as well as fake ones. Also the questions of ethics regarding the usage of such models, especially in media-related areas.
To sum up, False image recognition remains an immature and swiftly developing space of research with machine learning. The literature that exists poses the potential for using different machine learning techniques to detect fake images such as CNNs, GANs and RNNs. However, more research is required to create stronger and efficient models; not to mention ethical concerns related to their use.

III. PROPOSED METHODOLOGY

In this article, we will first introduce the concept of fake image detection using machine learning techniques. Here's a possible approach for constructing such model:

Get a huge dataset of real and fake images. The dataset should be diverse, covering different kinds of fake images — whether they are manipulated or deep faked. It also ensure the dataset must be balanced is balanced, i.e., the number of real and fake images are roughly the same[19]. Load and process images to a fixed size and colour space. For example, edges texture, colour histogram is helpful features that could be extracted[20]. Determine which machine learning algorithm you should use in detecting fake images like the Convolutional Neural Network (CNNs) that are especially effective for image classification. If you are using deep learning, try different architectures and hyperparameters to get the optimal model for the task[21].

opt for the selected model trained over the prepared dataset up until an appropriate loss function (binary cross-entropy, etc.) is utilized to enhance its performance. Regularization such as dropout regularization for example can be used to avoid the over fitting SQL[21]. Measure the accuracy, precision, recall and F1-score of the trained model on a held-out test dataset. If things don't going well, please try adjusting the model architecture or retrain using different hyperparameters[23].

When the model is determined to perform efficiently, deploy it in a production environment where it can make real-time predictions on the images before complete object detection. It is vital to track the model's metrics constantly and if it drops, retrain[24]. In conclusion, the propose methodology for machine learning pattern involves a series of steps: collecting and preprocessing data like text, audio or visual material from various sources; choosing an appropriate approach; training and evaluating it. Following the this steps you will have a good working machine learning model for fake image detection.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026| DOI:10.15680/IJMRSET.2026.0904034

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

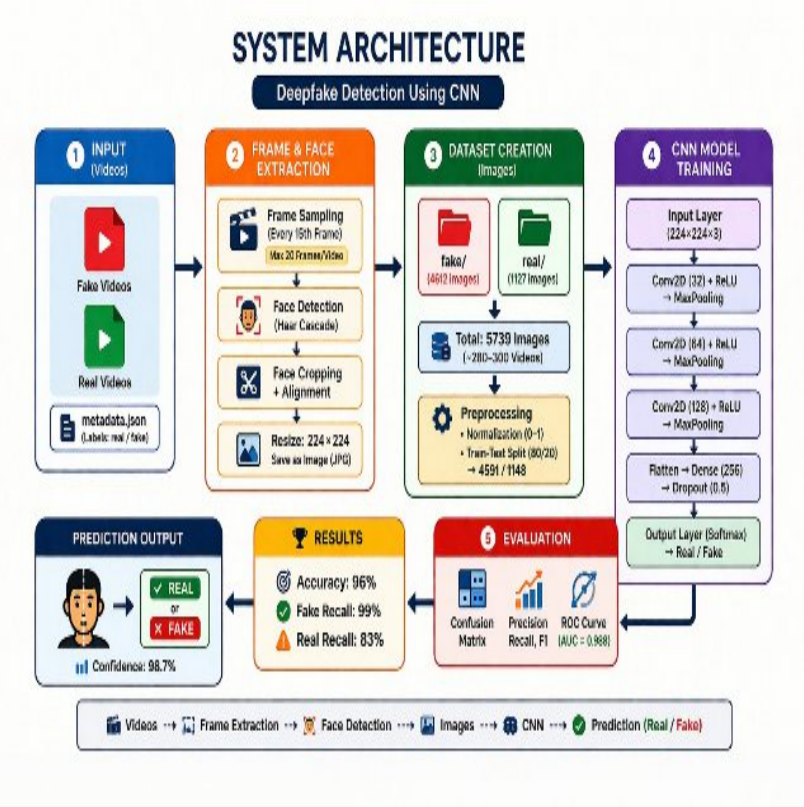
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. SYSTEM ARCHITECTURE

The proposed deepfake detection system follows a structured pipeline that converts videos into facial images and classifies them using Convolutional Neural Network (CNN). Initially, the system receives real and fake videos along with metadata file containing their labels. Since deep learning models work effectively with images, the videos are first processed through a frame sampling stage where selected frames are extracted at regular intervals to reduce and computational cost.

Each sampled frame is then passed through a face detection module using Haar Cascade method to identify the facial region. The detected faces are cropped, aligned and resized to a fixed resolution of 224*224 pixels to maintain uniformity. These processed face images are organized into two classes, real and fake, forming the dataset. Before training, the images undergo preprocessing, including normalization and train-split to ensure unbiased model evaluation.



Then, they use the processed dataset to train a CNN model used in a specific visual artifact that often appears on deepfake content like blending inconsistency and unreasonable sorts. The network is made up of many convolutional layers and pooling to learn hierarchy features, flatten layer so the feature maps were changed into vector and fully connected layers such as dropout helps reduce overfitting. The last output layer predicts the output as real face image or fake face based on binary classification. Once trained, the model is evaluated depending on various performance metrices like accuracy, precision, recall, F1-score confusion matrix and ROC curve. At inference time, the same pipeline is also used on new videos and the system predicts if a video is real or deepfake with confidence score thereby making the architecture applicable to real world deepfake detection usage.

V. SYSTEM WORKING

The algorithm works as an automatic pipeline which, using a CNN trained model, identifies if a video is real or deepfake based on facial characteristics. It all starts when users input a video. As videos are made by frames so system extracts the frame at regular intervals to reduces redundancies and capture important facial variations like expressions,

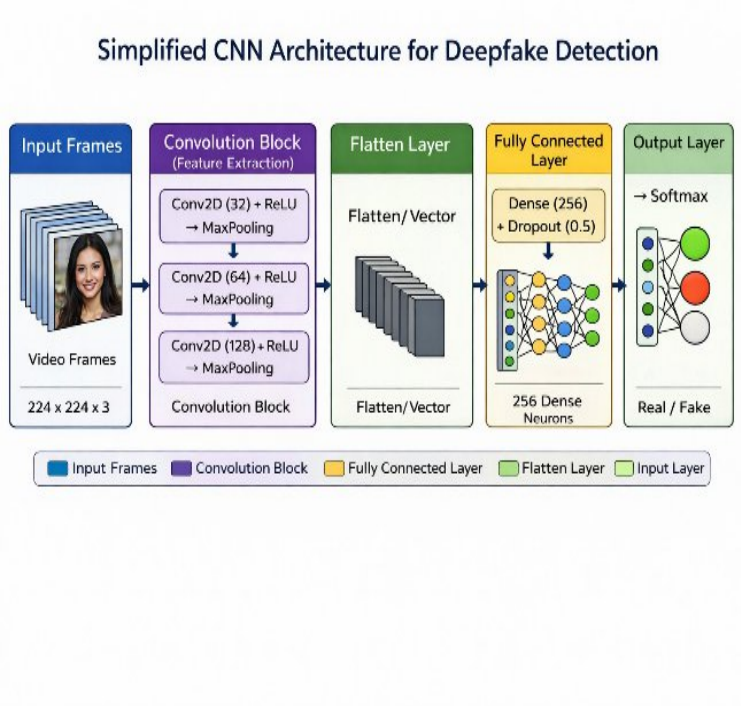
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

light and head movements. This allows the model to access representative visual information without evaluating every frame, which would be computationally costly.

\



Once frames have been extracted, each of these frames is run through a face detection module to find and crop the facial region from the background. Finally, the detected faces are cropped and aligned so that face is centred and consistent across all images. Talk about an important step, deepfake artifacts tend to only show up around the face area so by removing background noise helps the model focus on features that matter. The extracted faces are resized to a constant resolution and normalized, by scaling all pixels values in the same range, what allows stable and efficient learning.

These face images are then fed to the trained Convolutional Neural Network. Now, the CNN learns hierarchical features from images automatically. In the initial layers, the network identifies basic patterns like edges and textures, combined edges, lighting discrepancies and compression artifacts that are hard for humans to see. The detectors extract these learned features, followed with fully connected layers responsible for the final decision.
After analyzed the features, the output layer will classify whether input is real or fake. The system also provides a confidence score reflecting how confident the model is in its predication. For a new video, this whole pipeline is performed automatically which leads from extraction to face detection and CNN classification ultimately resulting in a final predication that helps us decide if the video is real or fake.

VI. RESULTS

In deep fake detection, we use machine learning to train our model using a dataset containing both real and synthetic images and then feed new images into our trained model for classification. The success of this process relies on the quality of the training data, how complex the model is

© 2026 IJMRSET | Volume 9, Issue 4, April 2026| DOI:10.15680/IJMRSET.2026.0904034

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

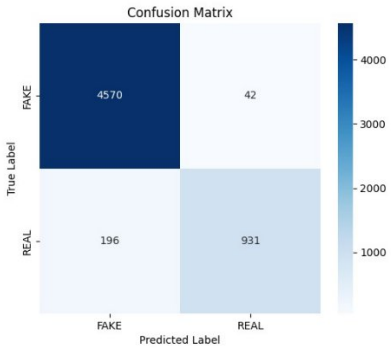
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

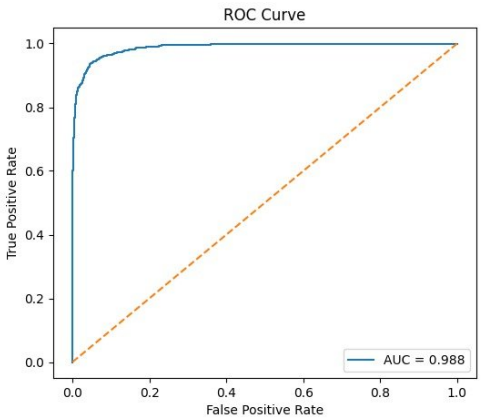
PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

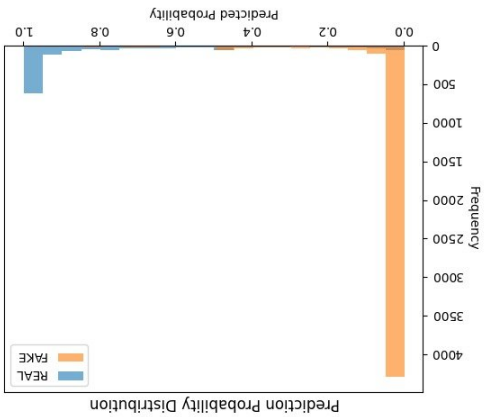
	precision	recall	f1-score	support
FAKE	0.96	0.99	0.97	4612
REAL	0.96	0.83	0.89	1127
accuracy			0.96	5739
macro avg	0.96	0.91	0.93	5739
weighted avg	0.96	0.96	0.96	5739

and the methods used to detect them. Machine learning algorithms are good at spotting simple manipulations like those involving resizing or cropping.



Though, it is far more difficult to detect sophisticated forgeries like deep fakes due to their complexity and realism.



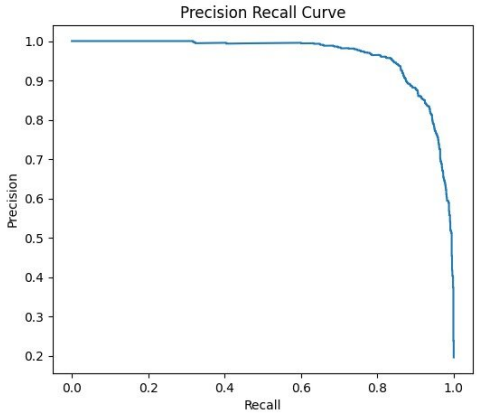


Machine learning is an effective instrument for detecting counterfeit images. ” It’s critical to note that no means of detection is perfect.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Verifying the authenticity of an image still depends on human judgement and expertise, since no amount of machine learning alone is likely to correctly catch all types of fakes.

VII. CONCLUSION

We proposed a methodology for differentiating between real and deep fake videos, along with the reliability of our model predications. We process one video frame per second through our model and obtain high accuracy. Using the CNN model, we try to extract frame-level features and perform classification whether the video is authentic or not. Our model converts the video into a stream of frames for predictions.

VIII. FUTURE SCOPE

Moreover, this dataset and associated model can be leveraged to integrate detection of deepfake audio, video or lop-synced videos.

This model can be used to detect audio signals that are artificially generated.

REFERENCES

[1] ARossler, DCozzolino, L Verdoliva, “FaceForensics++: Learning to detect Manipulated Facial Images” in arXiv:1901.08971
[2] Dataset: https://drive.google.com/drive/folders/1AwQ8adyGquf8uMeuHJk-HjsKfnZ3D6V?usp=sharing
[3] YLi,XYang, Pu Sun ”Celeb-DF:A Large -scale Challenging Dataset for Deep Deep Fake Forensices” in arXiv:1909.12962
[4] Deepfake examples that terrified and amused the internet:https://www.creativeblog.com/features/deepfakes-examples
[5] G. Antipov, M. Baccouche, and j-L. Dugelay. Face aging with conditional generative adversaria networks. arXiv:1702.01983, Feb 2017
[6] J.Thies et al.Face2Face: Real-time face capture and reenactment of rgb videos. Proceedings of the IEEE Conference on Computer vision and Pattern Recognition, pages 2387-2395, June 2016. Las Vegas,NV.
[7] Deepfakes, Revenge Porn, And THE Impct On Women: https://www.forbes.com/siteschenxiwang/2019/11/01/deepfakes-revenge-porn-and the -impact-on-women/
[8] The rise of the deep fake and threat to democracy https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy
[9] Deep Fake using GAN Tianiang shen Ruian liuju Bai Zheng Li Report16.pdf(uscd.edu)

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com