IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

Detecting Synthesized Faces Using Deep CNN Architecture

Prof. N. B.Vikhe[1], Saba Shaikh[2], Shweta Avhad[3], Shalini Shelke[4], Tanuja Adhav[5]
Department of Computer Engineering[1]
Department of Computer Science & Design[2-5]
Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar

Abstract: *Recent improvements to deepfake technology have raised worries about incorrect information, security flaws, and identity theft. This study presents a hybrid deep learning model that includes convolutional neural networks (CNN) to correctly identify phoney faces. The technique employs CNN and transformer-based architectures to detect anomalies in generated images across space and time. We use feature extraction and attention approaches to enhance detection robustness against hostile attacks. When compared to benchmark datasets, the proposed model outperforms previous approaches. Real-time deepfake detection is critical for digital security and media integrity, and our technology offers a reliable and extendable solution*

Keywords: CNN, Deep learning, Feature extraction, classification, Image Processing

I. INTRODUCTION

The benefits of GAN and ResNet architectures are merged in this study to produce a novel hybrid deep learning model for identifying phoney faces. The suggested model provides a breakthrough way for distinguishing between real and fake faces by combining the distinctiveness of GANs with the discriminative power of RESNET. The hybrid model's performance is measured against popular pre-trained models such as VGG16 and RESNET 50 [1].

The primary goal of deepfakes is to disseminate false information, as advances in computer vision and natural language processing have resulted in the development of deep learning (DL) and machine learning (ML) models. Humans can now be represented by phoney streams of sound, video, and photographs, or their visual and auditory characteristics can be exploited to adapt to different situations, thanks to advancements in generative adversarial networks (GANs).

Deepfakes are high-quality videos modified to appear authentic. To address the issues raised by Deepfake, several techniques have been proposed in the literature. In this study, we conducted a systematic literature review (SLR) to present a current overview of deepfake detection research, synthesising 112 relevant publications published between 2018 and 2020 using a variety of methodologies [3].

Inception-ResNet-v2 and XceptionNet are two powerful deep learning-based CNN architectures utilised in this study to detect deepfakes. In addition to being more accurate and efficient than earlier methods, the proposed methodology offers the most value for the given temporal and spatial complexity [4].

While there are numerous applications for temporal anomalies in edited movies, most studies have concentrated on spatial issues. We offer a hybrid deep learning system that efficiently distinguishes between real and false videos by incorporating temporal, spectral, and spatial data. We demonstrate how the discrete cosine transform, which looks at the spectral properties of each frame, can be used to improve deep fake detection[5].

A. Background Overview

Recognisingphoney faces has become more critical in maintaining digital security and trust, especially when dealing with deepfakes and AI-manipulated pictures. Differentiating between artificial intelligence-generated false faces has become more challenging since the development of generative models like GANs (Generative Adversarial Networks). This issue could be addressed using hybrid deep learning models based on convolutional neural networks (CNNs). CNNs are useful for depiction analysis because they can automatically learn spatial hierarchies from data. CNNs use

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

fine-grained information like texture, illumination, and pixel patterns to tell the difference between real and fraudulent photos.

Deepfakes introduce sophisticated, nuanced artefacts that CNNs may struggle to detect. Transformers, attention processes, and recurrent neural networks (RNNs) are examples of hybrid systems that employ advanced techniques to enhance CNN performance. These are effective for identifying temporal or spatial correlations and focus on the most significant data. RNNs, for example, may detect frame anomalies in deepfake films, while attention layers allow the model to highlight suspicious spots in images.

II. RELATED WORK

Yogesh Patel [1] has carried out a number of research to investigate how social media has enhanced users' access to multimedia material in smart communities. Recent developments in computer vision and natural language processing have led to the development of deep learning (DL) and machine learning (ML) models. Generative ad networks (GAN) enable the imitation of audio, visual, or video streams and can adapt to changing contexts according to user vision and sound preferences. Deepfakes are frequently used on social media platforms to spread misleading information and content that harms a person's or company's reputation. Several recent studies have focused on the production and detection of deepfake audio and video streams.

In 2022, Dr. ShahelRana [2] explained. Over the last two decades, tremendous breakthroughs in artificial intelligence, machine learning, and deep learning have resulted in the creation of new methodologies and tools for manipulating multimedia. While technology has primarily been used for positive purposes such as entertainment and education, dishonest people have abused it for harmful or unlawful objectives. For example, extremely convincingly fake audio, video, or image content that appears authentic has been used to disseminate disinformation and propaganda, cause political unrest and enmity, and even threaten and blackmail individuals. These realistic-looking, meticulously created, and recently widely distributed altered videos are referred to as "deep fakes."

John K. Lewis and colleagues conducted a study in 2020 [3]. Digital media authentication is becoming a more crucial part of modern life. The detection of synthetic media has become increasingly challenging since the emergence of Generative Adversarial Networks (GANs). Deepfakes, or films including altered versions of actual people's voices and faces, endanger online privacy and trust. Deepfakes can be used to disseminate misinformation, achieve political objectives, and harm well-known individuals' reputations. Despite the limitations of deepfakes, many people are unable to distinguish between authentic and fraudulent images and videos. As a result, automated mechanisms for detecting whether digital content has been accepted must be implemented.

Michael Tsang and colleagues conducted a study in 2021 [4]. This paper introduces a novel human-centred approach to identifying facial photo fraud that employs dynamic prototypes for visual explanation. The vast majority of effective deepfake detection systems use black-box models to evaluate movies frame by frame while ignoring temporal anomalies. Temporal irregularities are essential for human supervisors to identify and comprehend deepfake films. Our Dynamic Prototype Network (DPNet) uses dynamic representations to explain temporal artefacts in deepfakes.

besides others. Kaihan Lin [5] conducted research in 2021. When used inappropriately, deepfakes can be harmful to both individuals and society. Researchers have looked into ways to detect deepfakes and lessen the damage they inflict. Despite enormous progress, researchers have been unable to fully explain deepfake manufacturing and detection methods due to competing goals. This paper provides a scientific classification of existing practices and data sets, along with a detailed evaluation of deepfake development and detection methods. Our findings will be useful in detecting deepfakes in the future.

Challenges in deep Fake detection

Several research studies have proven that it is difficult to detect deepfakes. The article "Hybrid Deep Learning Model Based on GAN and RESNET for Detecting Fake Faces" tackles topics such as generalisation to veiled deepfakes and adversarial robustness. The essay "Deepfake Generation and Detection: A Case Study and Challenges" (repeated) investigates how rapid advances in deepfake algorithms have rendered detection techniques useless. "Deepfake Video

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning" was created to detect deviations caused by changes in illumination, resolution, and compression artefacts. The main concerns include adversarial attacks, real-time detection efficacy, dataset bias, and responding to new deepfake approaches.

III. EXISTING SYSTEM

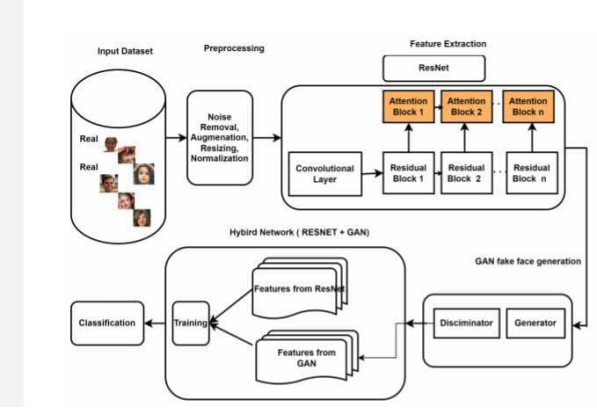

Fig : Existing System of Deep Fake Detection

The diagram demonstrates the current deepfake detection system, which uses ResNet and GAN structures to identify fake faces. The technique starts with a set of inputs that includes both genuine and synthetic faces. Scaling, augmentation, normalisation, and noise reduction are all examples of preprocessing procedures that improve data quality. ResNet is utilised to extract features, and a convolutional layer captures fine-grained facial features before a series of residual and attention blocks. A hybrid adversarial training model using ResNet and GAN features makes fake faces. The GAN's discriminator makes it easier to discriminate between genuine and fake photos.

IV. SYSTEM ARCHITECTURE

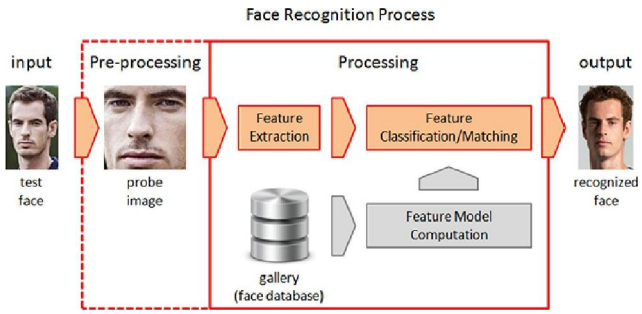

Fig. System Architecture

The CNN-based deepfake facial recognition system is built through several processes, including input collecting, preprocessing, feature extraction, classification, and output synthesis. The initial phase in the procedure is to compile a collection of images, including both actual and phoney facial shots. Scaling, normalisation, and image augmentation are used as preprocessing approaches to improve the model's performance.

A convolutional neural network (CNN) retrieves key facial characteristics such as texture, edges, and spatial patterns. Following data collection, the classification system distinguishes between real and fake faces using previously learnt patterns. The classifier employs sigmoid or softmax activation to determine if a picture is real or not based on its confidence score.

The image is then identified as "real" or "fake." Because of its speed and accuracy in deepfake identification, this technology is suitable for use in digital forensics and media authentication.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

V. ALGORITHM

Convolutional neural networks are a type of deep learning algorithm that is specifically intended to analyse and interpret visual data. These networks are often referred to as "CNN" networks. CNNs are widely utilised in a number of computer vision applications, including image classification, object recognition, and segmentation. Convolutional Neural Networks, often known as CNN or ConvNet, can detect and categorise images. This is how supervised learning is used with unexpected feedforward neural circuits. Convolutional Neural Networks can quickly establish the relationship between unprocessed pixel input and class labels via end-to-end learning.

This model was trained using a vast amount of data. This approach has the potential to increase the accuracy of traditional clinical picture classification.

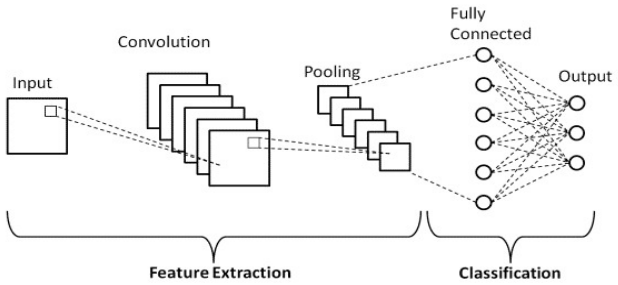

Fig. CNN Architecture

Each pixel contains eight bits (a byte). The network does not learn colours. Because computers can only understand 1s and 0s, colour values are transferred across the network in binary. A feature detector, a feature map, and an input image constitute the Convolution Operation's three components. Feature maps are used to minimise the size of the input images.

Convolutional Neural Networks have following layers:

Convolution: Convolution uses filters (kernels) to extract important aspects from an input image, such as edges, textures, and patterns.

ReLULayer :This activation function brings nonlinearity into the model by turning all negative values to zero while maintaining positive values constant.

Pooling: Pooling, also known as maximum pooling, reduces the spatial dimensions of feature maps while retaining important data.

Fully Connected: The final layer flattens the collected features before categorising the image as melanoma, benign, or other skin disorders.

VI. RESULT AND DISCUSSION

Finally, the deepfake detection system uses a hybrid model of ResNet and GAN architectures to recognise fake facial images. After noise reduction, augmentation, and normalisation, the system extracts deep features from input pictures using ResNet's convolutional and attention blocks. The GAN model increases detection and resistance by generating adversarial false photos. Increasing model generality to hide deepfake tactics, minimising dataset bias by enriching training data, and using continuous learning to counteract adversarial attacks are among the topics discussed. The method uses multimodal analysis to avoid false positives. In practice, the final solution improves accuracy, strengthens particular deepfake approaches, and includes a sophisticated classification mechanism for real-time detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

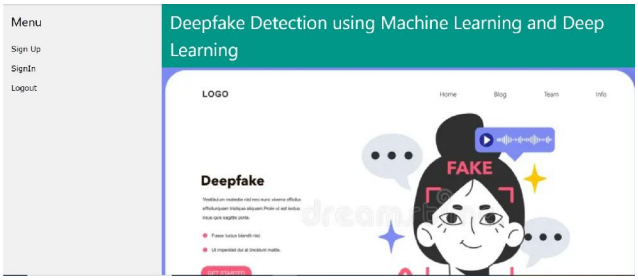

Fig : Home Page

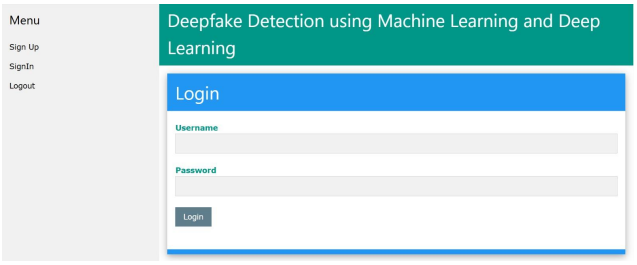

Fig : Login Page

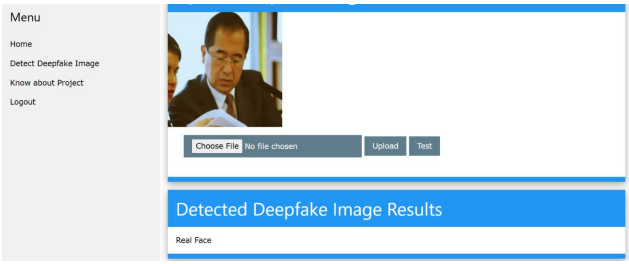

Fig : Real Face Image

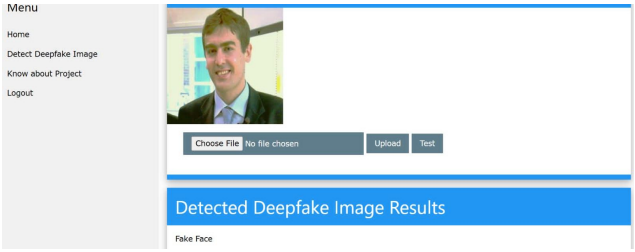

Fig : Fake Face Image

Result Graph :

The dataset size graph shows a balanced distribution of actual and created deepfake images, ensuring fair model training. To show the model's learning process, the accuracy graph plots training performance over epochs, beginning with a considerable increase and then stabilising. A large decrease at the start of the loss graph, which shows the loss decreasing with time, suggests rapid convergence, but it could also indicate overfitting. According to the training log, an average accuracy of 54% indicates that feature extraction and classification may require additional effort. These findings demonstrate that the model can recognisedeepfakes, but they also highlight the need for more optimisation to improve generality and accuracy, such as changing hyperparameters or expanding the dataset.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

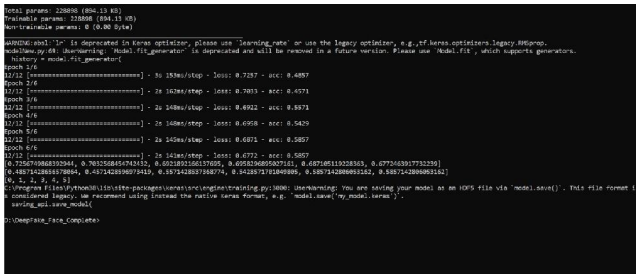

Fig : Model Training

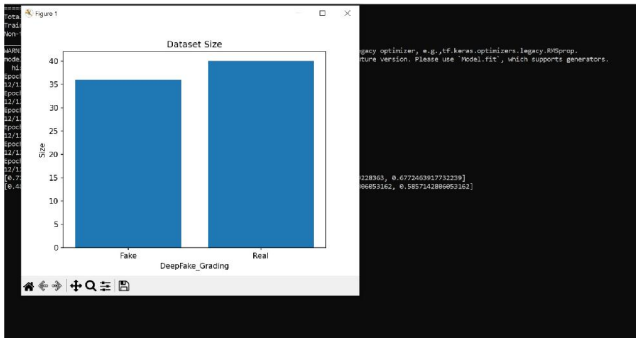

Fig :Dattset Size

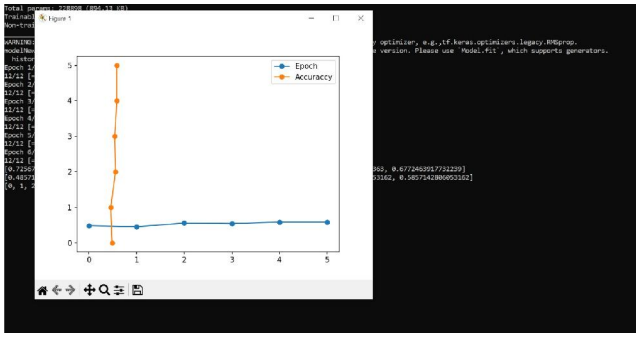

Fig : Accuracy Graph

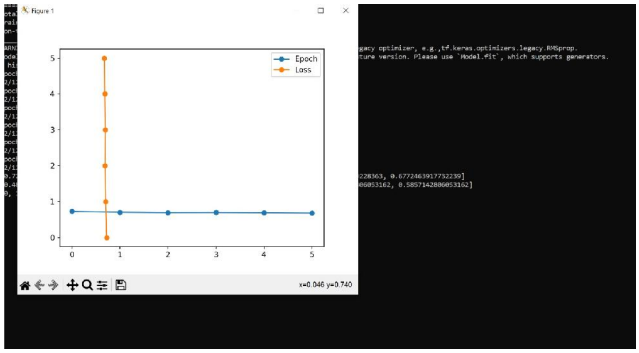

Fig : Loss Graph

VI. CONCLUSION

To reduce criminal activities and misleading intake by unauthorised users or those who utilise phones, masks, or posters to commit facial spoofing attacks. It accurately mimics a wide range of techniques, including photo, video, and mask attacks. Because of its fast processing time, the proposed technique may be capable of recognising a huge number of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, June 2025
Impact Factor: 7.67

faces, making it perfect for quick suspect searches. Despite improvements in accuracy and robustness, these models continue to have high processing costs, complex design, data dependencies, and attack vulnerabilities. Large, diversified datasets are required to generalise to the ever-changing types of artificial facial systems.

Future deepfake detection research should concentrate on increasing real-time detection efficiency, incorporating multimodal analysis, and enhancing model robustness against novel deepfake tactics. Using larger and more diverse datasets, optimising adversarial training, and incorporating blockchain for authentication all help to improve detection accuracy and dependability in real-world scenarios.

REFERENCES

[1]. S. Safwat, A. Mahmoud, I. EldesoukyFattoh and F. Ali, "Hybrid Deep Learning Model Based on GAN and RESNET for Detecting Fake Faces," in IEEE Access, vol. 12, pp. 86391-86402, 2024, doi: 10.1109/ACCESS.2024.3416910.

[2]. Patel, Yogesh, SudeepTanwar, Rajesh Gupta, Pronaya Bhattacharya, Innocent Ewean Davidson, RoyiNyameko, SrinivasAluvala, and VrinceVimal. "Deepfake generation and detection: Case study and challenges." IEEE Access (2023).

[3]. Rana, MdShohel, Mohammad NurNobi, BeddhuMurali, and Andrew H. Sung. ”Deepfake detection: A systematic literature review.” IEEE access 10 (2022): 25494-25513.

[4]. Trinh, Loc, Michael Tsang, SirishaRambhatla, and Yan Liu. ”Interpretable and trustworthy deepfake detection via dynamic prototypes.” In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1973-1983.

[5]. S. P and S. Sk, "DeepFake Creation and Detection:A Survey," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 584-588, doi: 10.1109/ICIRCA51532.2021.9544522.

[6]. D. Pan, L. Sun, R. Wang, X. Zhang, and R. O. Sinnott, ”Deep fake detection through deep learning,” *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, Leicester, UK, 2020, pp. 134–143, doi: 10.1109/BDCAT50828.2020.00001.

[7]. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, ”DeepFake detection for human face images and videos: A survey,” *IEEE Access*, vol. 10, pp. 18757–18775, 2022, doi: 10.1109/ACCESS.2022.315118.

[8]. S. R. B. R, P. Kumar Pareek, B. S., and G. G., ”Deepfake video detection system using deep neural networks,” *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*, Raichur, India, 2023, pp. 1–6, doi: 10.1109/ICICACS57338.2023.10099618.

[9]. M. S. Rana, B. Murali, and A. H. Sung, ”Deepfake detection using machine learning algorithms,” *2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Niigata, Japan, 2021, pp. 458–463, doi: 10.1109/IIAI AAI53430.2021.00079.

[10]. P. Theerthagiri and G. B. Nagaladinne, ”Deepfake face detection using deep InceptionNet learning algorithm,” *2023 IEEE International Students’ Conference on Electrical, Electronics, and Computer Science (SCEECS)*, 2023