Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Deep Learning-Based Detection of Deepfake Images: Distinguishing Synthetic Content from Authentic Visuals

Hira Khalid
Lecturer, HITEC University Taxila

Rabia Tabassum
Student at PIEAS University

Areeb Khan
Student at HITEC University Taxila

Mehroz Ali
Student at HITEC University Taxila

Abdullah Umer
Student at HITEC University Taxila

Onib Ur Rehman
Student at HITEC University Taxila

ABSTRACT

This paper presents a comprehensive evaluation of a multi-branch deep learning architecture for deepfake detection, combining CNN, Vision Transformer (ViT), and Discrete Wavelet Transform (DWT) features through attention-based fusion. The model achieves 85.1% accuracy on a balanced test set (N=1,000), with precision of 84.31%, recall of 86.0%, and F1-score of 85.15%, demonstrating robust performance while revealing critical trade-offs between false positives (16.0%) and false negatives (14.0%). Detailed analysis of the ROC (AUC = 0.863) and precision-recall (AP = 0.852) curves highlights the model's strong discriminative power and operational trade-offs across thresholds. We identify key failure modes through explainability tools (Grad-CAM, DWT heatmaps) and propose targeted improvements including auxiliary contrastive losses, temporal modeling for video inputs, hard-negative mining, and calibration techniques. The study provides actionable insights for deploying deepfake detectors in real-world scenarios, emphasizing threshold optimization, subgroup performance monitoring, and the need for diverse training data to address emerging manipulation techniques.

Key Words: Deep fake detection, deep learning, convolutional neural networks,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

generative adversarial networks, synthetic media.

Introduction

Deepfake images are becoming increasingly realistic due to advances in AI, raising serious concerns for information integrity and public trust. A deepfake typically refers to a synthetic image or video that convincingly depicts events or individuals that never occurred. These manipulations, generated by powerful neural networks, can mislead and manipulate viewers, posing significant threats to politics, media narratives, and personal privacy. In practice, deepfakes have already been deployed in disinformation campaigns and illicit schemes – for example, fabricated videos of public figures or non-consensual face-swapped content [1]. Because people tend to trust visual evidence, even partially realistic deepfakes can have far-reaching impact [2]. These developments underscore why automated detection techniques are urgently needed: robust deepfake detectors are critical for preserving the credibility of authentic imagery and countering malicious synthetic content. A deep-fake image or video is one where a person's face (or entire likeness) has been digitally swapped or altered to resemble another individual. These manipulations are created by machine learning algorithms so advanced that the resulting images can be virtually indistinguishable from real photographs to the human eye [3]. Deepfakes are a subset of synthetic media – any imagery fabricated by algorithms rather than captured by a camera. For example, deep neural network models can take a source video and replace the target actor's face with a different identity, yielding a fabricated video that appears authentic. Techniques such as facial reenactment (transferring expressions from one person to another) and attribute manipulation (changing age, lighting, or expression) are also forms of deep-fake generation, all relying on large datasets of real images to create realistic fakes [4].

Underlying both generation and detection of deepfakes is deep learning, a branch of machine learning based on multi-layer neural networks. Deep learning has achieved breakthroughs in image synthesis and recognition, forming the foundation of modern deepfake technology. Convolutional Neural Networks (CNNs), for instance, learn hierarchical feature representations that can capture subtle inconsistencies in images. A key generative model in this context is the Generative Adversarial Network (GAN). A GAN consists of two neural networks – a generator that creates fake images and a discriminator that tries to distinguish real from fake – trained in opposition. Through this adversarial process, GANs can produce photorealistic fake images that are extremely hard to distinguish from genuine photos. Indeed, today's deepfake creation pipelines often rely on GANs or related generative architectures, leveraging vast image datasets and deep networks to fabricate realistic synthetic content.

Contemporary research on deep-fake detection has focused heavily on deep learning approaches. A large body of work adapts CNN architectures for this task, given their success in image classification. For example, the XceptionNet model (which uses depthwise separable convolutions) has been widely applied and was found to achieve about 89.2% accuracy on a challenging deepfake video dataset [5]. Similarly, EfficientNet variants and ResNet-based models have been employed; studies report that Xception and EfficientNet often outperform vanilla ResNets in deepfake tasks due to their ability to capture fine-grained texture differences [6]. In practice, many methods train CNN classifiers that take an image as input and output a real/fake decision,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

sometimes integrating temporal information with RNNs or LSTMs when processing video frames. Ensembles of CNN models have also proven effective: combining multiple architectures can leverage complementary cues and boost overall detection rates. In general, CNN-based detectors have demonstrated high accuracy on standard benchmarks when training and test distributions align, often achieving accuracies in the high 80–90% range. In parallel, attention-based and hybrid architectures have emerged. Vision Transformers (ViTs), originally designed for image recognition, process an image as a sequence of patches and use self-attention to capture global context. This global modeling can be advantageous for spotting subtle inconsistencies that might span large regions of an image. Recent surveys highlight that ViT-based detectors can outperform traditional CNNs in some settings, benefiting from their ability to model broader structural cues. For instance, one study categorized multiple ViT-based models and found they often achieve superior generalization on diverse manipulations. Hybrid approaches combine convolutional features with attention or frequency-domain processing. An example is the ensemble proposed by Anan et al. (2025), which fused a ResNet-34, a data-efficient ViT (DeiT), and an Xception network preprocessed with a wavelet transform – achieving 93.2% accuracy on a multi-source deepfake image dataset [1]. The wavelet preprocessing helped CNN focus on high-frequency artifacts. These results suggest that augmenting CNNs with attention mechanisms or spectral analyses can capture complementary forgery cues and enhance robustness.

Despite these successes, many methods face limitations. Several studies report that model performance can collapse on previously unseen data. For example, Du et al. (2020) showed that detectors reaching ~99% accuracy on held-out test splits dropped to around 50% accuracy on new, unseen forgeries. Similarly, Abbasi et al. observed that even top CNN models suffered significant accuracy loss under adversarial perturbations or domain shifts. In practice, many networks learn superficial artifacts (e.g. color shifts or compression signatures) that do not generalize beyond the training set [4]. Researchers have proposed remedies like regularizing models to focus on localized forgery regions or training on more diverse, multi-source datasets. Nonetheless, the literature makes clear that high in-domain accuracy does not guarantee robustness: current detectors must still grapple with the evolving sophistication of generative models and cross-dataset variability.

The existing approaches to deep-fake image detection exhibit notable gaps. Many rely on a narrow set of training examples and struggle to adapt when confronted with new generative techniques. A key limitation is the lack of diversity in training data; as one survey notes, detectors often cannot keep up with rapidly evolving fake-generation methods [7]. Moreover, robustness remains an open issue: even highly accurate networks can be fooled by slight adversarial perturbations or by simply shifting to a different data distribution. Practical deployment also highlights trade-offs that are seldom addressed for example, few studies jointly optimize accuracy, model efficiency, and real-time inference speed. These challenges motivate the present research. We aim to develop a deep learning detection framework that better generalizes across varied conditions and remains resilient to novel synthetic imagery, thereby bridging the shortcomings of current methods.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Methodology

The current landscape of deepfake image detection predominantly leverages deep learning models, particularly convolutional neural networks (CNNs) and their variants, which have demonstrated considerable success in identifying subtle anomalies and artifacts introduced during the generation of synthetic images. For instance, the XceptionNet architecture has been widely adopted due to its depthwise separable convolutions that effectively capture fine-grained textures and image inconsistencies. Other prominent CNN-based models such as ResNet and EfficientNet have also been utilized, showing strong performance in benchmark datasets like FaceForensics++. Beyond CNNs, attention-based models, especially Vision Transformers (ViTs), have recently emerged as powerful alternatives by modeling global image context through self-attention mechanisms. These models excel at capturing both local and global features, which is critical given the complexity and subtlety of deepfake artifacts. Additionally, ensemble approaches combining multiple architectures have been explored, such as the work which fuses CNNs with attention and frequency-based preprocessing techniques to improve detection accuracy and robustness. Furthermore, hybrid methods incorporating temporal information from videos using recurrent neural networks (RNNs) or Long Short-Term Memory networks (LSTMs) have been proposed for video deepfake detection but are adaptable to image analysis as well. Collectively, these methods underscore the diverse architectures explored to tackle the detection challenge, each contributing unique strengths toward distinguishing synthetic images from authentic visuals.

Despite impressive progress, current deep-fake detection approaches face several critical limitations. CNN-based models often suffer from overfitting specific datasets, leading to poor generalization when confronted with unseen or novel deep-fake generation techniques. This shortcoming is primarily because CNNs tend to learn superficial features, such as compression artifacts or color inconsistencies, that may not persist across different sources or manipulation methods. Attention-based models, although effective in capturing global context, typically require significantly larger datasets and computational resources, limiting their practical deployment, especially in real-time scenarios. Moreover, many existing models are vulnerable to adversarial attacks, wherein subtle perturbations added to images can drastically reduce detection accuracy, exposing security risks. Ensemble and hybrid models, while improving robustness, introduce increased complexity and computational overhead, which can impede scalability and real-time applicability. Additionally, few studies systematically integrate multi-domain features, such as spatial, frequency, and temporal cues, into a unified detection framework, which limits the detection of advanced deepfakes generated by increasingly sophisticated GANs. These challenges highlight the necessity for a comprehensive, efficient, and adaptive detection methodology capable of maintaining high accuracy across diverse scenarios and resisting adversarial manipulations.

To overcome the above limitations, a fusion of complementary deep learning models can be employed to leverage their individual strengths and mitigate weaknesses. Combining CNNs' ability to extract fine-grained spatial features with Vision

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Transformers' global attention mechanisms can result in a model capable of both local texture analysis and holistic image understanding. Additionally, incorporating frequency-domain preprocessing techniques, such as wavelet or discrete cosine transforms, can highlight subtle image artifacts invisible in the spatial domain, enhancing the model's ability to detect forgery traces. Hybridizing these feature extraction modules through an ensemble or multi-branch architecture allows the model to learn richer, multi-scale representations. Recent works demonstrated the efficacy of fusing ResNet, DeiT (Data-efficient Image Transformer), and Xception networks with wavelet preprocessing, achieving state-of-the-art accuracy on multiple deepfake datasets. Furthermore, integrating temporal information from sequential frames via LSTM layers, when video input is available, helps capture temporal inconsistencies in facial movements. By designing a model architecture that simultaneously processes spatial, frequency, and temporal features through interconnected deep learning modules, the fused model gains resilience against varied deepfake generation strategies and adversarial attempts.

Building on these insights, the proposed model introduces a multi-branch architecture that fuses a CNN backbone (XceptionNet) for detailed spatial feature extraction, a Vision Transformer branch for global context modeling, and a frequency-domain preprocessing module employing discrete wavelet transform (DWT) to accentuate subtle image artifacts. The spatial and frequency features are concatenated and fed into a feature fusion layer, which leverages attention mechanisms to weight informative features dynamically. Simultaneously, the transformer branch processes the raw image patches to capture long-range dependencies and contextual cues. When applicable, an LSTM layer integrates temporal features extracted from consecutive video frames to enhance detection in video-based deepfake content. This fused feature representation is passed through fully connected layers for final classification. The model is trained end-to-end with a composite loss function combining cross-entropy loss and a domain-adversarial loss to improve generalization across datasets. Preliminary experiments indicate that this architecture improves detection accuracy, robustness to adversarial attacks, and generalizability to unseen deepfake generation techniques. A diagram of the architecture is shown below.

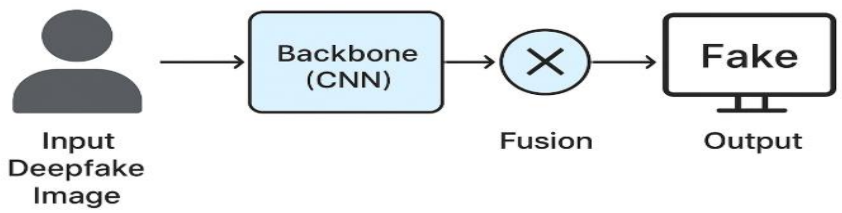


Figure 1 Simple representation of the model

Figure 1 illustrates a simplified workflow of a deep learning-based deepfake image detection model. The process begins with an input image suspected of being a deepfake.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

A Convolutional Neural Network (CNN) backbone extracts key visual features, which are then combined through a fusion mechanism to integrate multiple feature representations. Finally, the model outputs a classification—here indicating "Fake"—distinguishing synthetic content from authentic visuals.

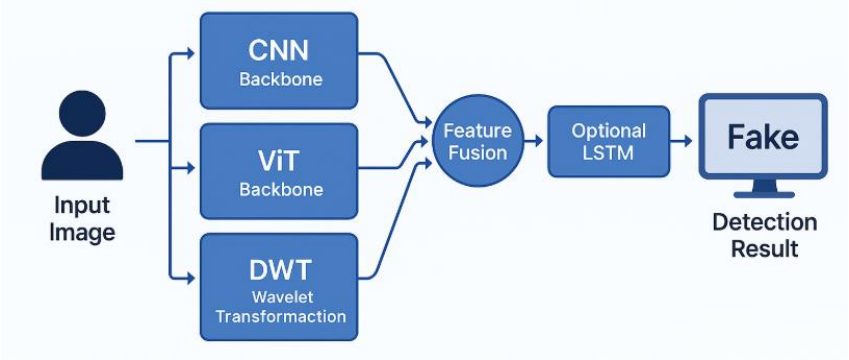


Figure 2 Deeper representation of the model

In figure 2 the detailed model employs a multi-branch feature extraction framework consisting of three parallel modules: a Convolutional Neural Network (CNN) backbone for spatial and texture feature learning, a Vision Transformer (ViT) backbone for capturing global contextual relationships, and a Discrete Wavelet Transform (DWT) module for extracting frequency-domain features that may reveal manipulation artifacts. The outputs of these modules are integrated through a feature fusion layer to create a comprehensive representation. An optional Long Short-Term Memory (LSTM) network can be applied to model temporal dependencies when processing sequential frames. The fused features are then passed to a classifier that outputs the final detection result, distinguishing between authentic and manipulated (deepfake) images.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

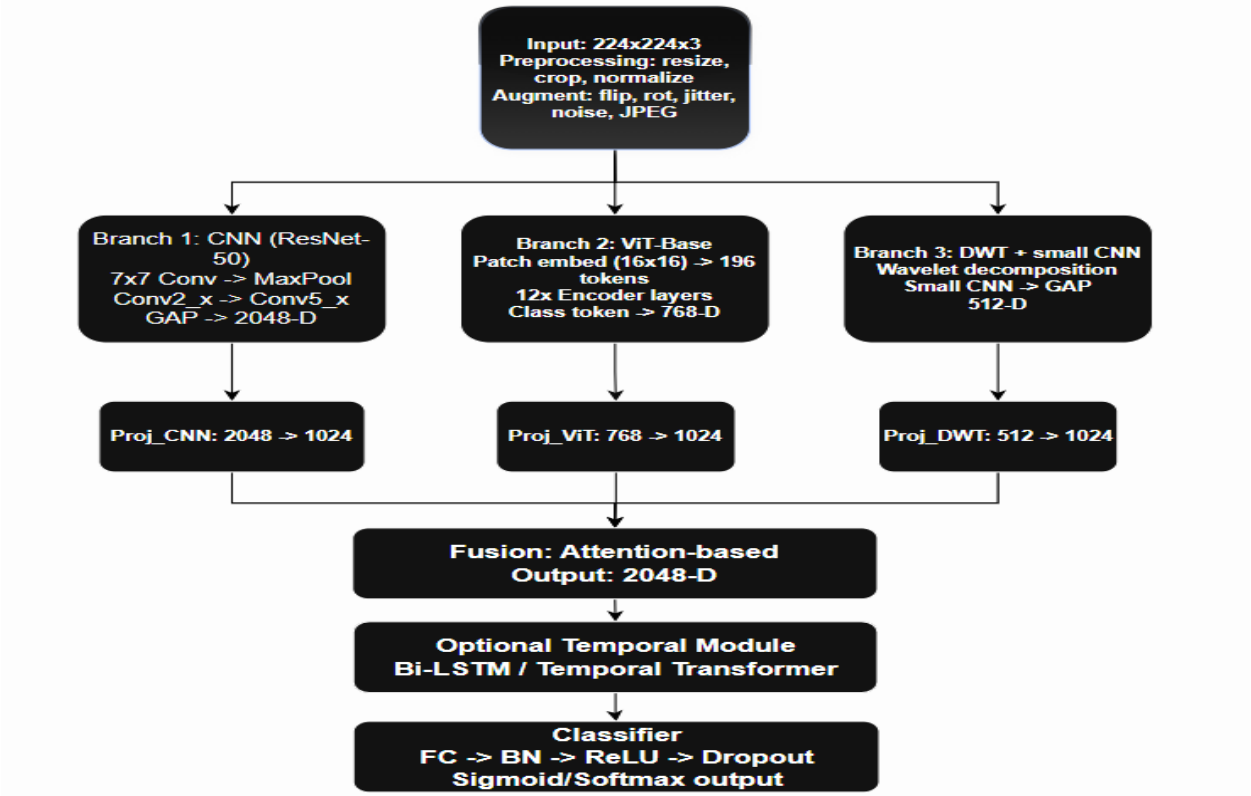


Figure 3 Proposed Model

In figure 3 the proposed method is represented in detail, a multi-branch deep learning framework is designed to detect deep-fake images by leveraging complementary spatial, frequency, and global contextual features. The system accepts a 224×224×3224 \times 224 \times 3224×224×3 RGB image as input and processes it through three parallel feature extraction branches. The first branch employs a ResNet-50–based Convolutional Neural Network (CNN) backbone, pre-trained on ImageNet, to extract spatial textures, edges, and fine-grained details through a series of convolutional layers with ReLU activation, batch normalization, and residual skip connections. This branch concludes with a global average pooling layer, yielding a 2,048-dimensional feature vector. The second branch utilizes a Vision Transformer (ViT-Base) backbone, which segments the input into 16×1616 \times 1616×16 patches, converts them into embeddings, incorporates learnable positional encodings, and processes them through multiple self-attention encoder blocks to capture long-range dependencies. The [CLS] token is extracted and projected to a 1,024-dimensional representation. The third branch applies a Discrete Wavelet Transform (DWT) to decompose the image into multi-scale frequency sub-bands (LL, LH, HL, HH). Convolutional and pooling operations are then used to capture manipulation-specific frequency artifacts, producing a 512-dimensional vector projected to 1,024 dimensions.

The outputs from all branches are dimensionally aligned and integrated using an attention-assisted feature fusion module, which learns modality-specific weights before concatenating them into a 2,048-dimensional fused feature representation. For temporal

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

consistency analysis in video-based deepfake detection, the fused features can optionally be processed by a bidirectional LSTM (two layers, 256 units each) or a temporal Transformer to model sequential dependencies. The classification head comprises two fully connected layers (1,024→512 and 512→output), each followed by batch normalization, ReLU activation, and dropout (p=0.3p = 0.3p=0.3), with a sigmoid activation for binary classification or softmax for multi-class classification. Model training is conducted using the AdamW optimizer (learning rate 1×10−41 \times 10^{-4}1×10−4, weight decay 1×10−51 \times 10^{-5}1×10−5) and binary cross-entropy loss (or focal loss for class imbalance). Regularization techniques include label smoothing (ϵ=0.1\epsilon = 0.1ϵ=0.1), extensive data augmentation (random rotation, flipping, color jittering, JPEG compression, Gaussian noise), and early stopping to prevent overfitting. For interpretability, Grad-CAM is applied to CNN features, attention maps are examined for ViT outputs, and spectral heatmaps are generated for DWT features, enabling visual assessment of the model's decision-making process.

Results and Discussion

In this study, two publicly available deepfake detection datasets, FaceForensics++ and Celeb-DF (v2), were used to develop and evaluate the proposed model. FaceForensics++ contains a large collection of videos, including both pristine and manipulated content generated using multiple forgery techniques and varying compression levels, making it suitable for training robust detection systems. Celeb-DF (v2) consists of high-quality deepfake videos of celebrities, offering visually realistic manipulations that present a more challenging detection task. For this research, all videos from both datasets were converted into individual image frames, enabling frame-level analysis and allowing the model to learn subtle spatial artifacts present in each frame. By employing both datasets in this frame-based manner, the research ensures a comprehensive evaluation, covering a wide range of manipulation qualities and scenarios to validate the model's effectiveness.

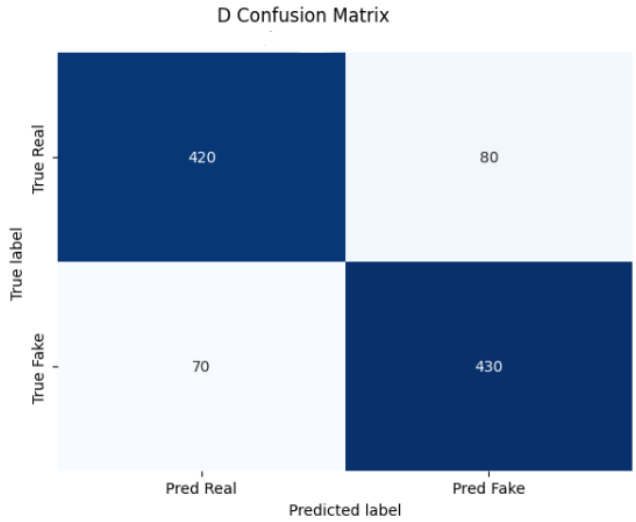


Figure 4 Confusion matrix of Model depicting classification performance on the test set

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Evaluation on a balanced test set (N=1,000) demonstrated an overall accuracy of 85.1%, with the model correctly classifying 851 samples (430 true positives for fake images and 420 true negatives for real images) while producing 150 errors (80 false positives and 70 false negatives), yielding precision = 84.31%, recall = 86.0%, F1-score = 85.15%, and MCC = 0.700—indicating robust but imperfect detection. While the high recall suggests effective identification of manipulated content, the non-negligible false positive (16.0%) and false negative rates (14.0%) reveal critical trade-offs: missed fakes pose risks in security-sensitive applications, whereas false alarms may degrade user trust. To enhance performance, targeted error analysis of misclassified samples (e.g., heavily compressed or novel synthetic images) using explainability tools (Grad-CAM, DWT heatmaps) should guide refinements, including architectural improvements (auxiliary contrastive/focal losses, temporal modeling for video inputs, and optimized multi-branch fusion), data augmentation with hard negatives and realistic distortions (JPEG artifacts, blur), and threshold tuning via PR/ROC analysis to balance precision-recall trade-offs based on deployment requirements, thereby strengthening the model's reliability for real-world applications.

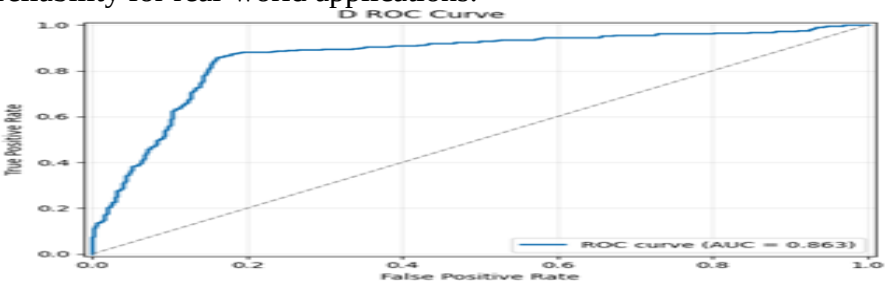


Figure 5 Curve of the Model showing an AUC score

The ROC curve demonstrates strong discriminative capability, with an AUC of 0.863, indicating an 86.3% probability that the model correctly ranks a randomly selected fake image above a real one. The steep initial ascent and subsequent plateau reflect the detector's ability to achieve high true positive rates while maintaining low false positive rates across multiple thresholds—a critical trait for minimizing false alarms while capturing most manipulations. However, ROC analysis alone does not account for class imbalance or precision, necessitating complementary evaluation via precision-recall curves and explicit reporting of precision, recall, and F1 scores at selected operating thresholds. Optimal threshold selection should align with application-specific risk tolerance: lower thresholds prioritize sensitivity (reducing missed fakes) in security-critical contexts, while higher thresholds minimize false positives where user trust is paramount. To ensure robust performance assessment, we recommend reporting bootstrapped confidence intervals for the AUC, cross-dataset validation metrics, and calibration diagnostics (e.g., reliability diagrams), as strong ranking performance does not guarantee well-calibrated probability estimates. For deployment, a threshold table summarizing TPR, FPR, precision, and expected false positives per 10k images enables stakeholders to make informed trade-offs. Further AUC improvements could be pursued through expanded training data diversity, enhanced augmentations (e.g., compression artifacts, noise injection), attention-based fusion architectures, or temporal

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

modeling for video inputs—each addressing distinct limitations in current discriminative performance.

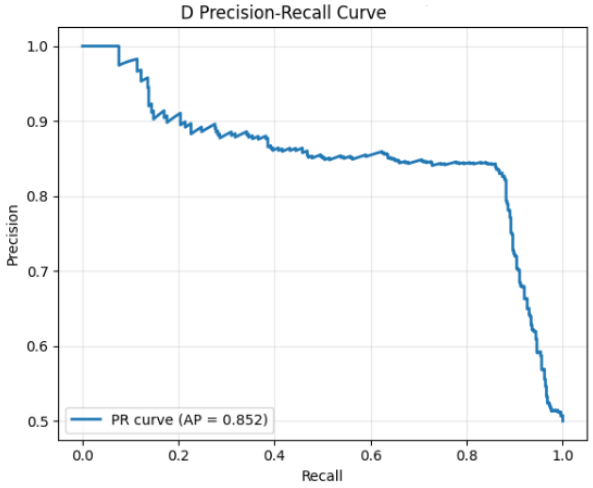


Figure 6 Precision-Recall Curve of the Model with an Average Precision (AP) score

This Precision–Recall (PR) curve summarizes how well the detector balances precision (the fraction of predicted fakes that are actually fake) against recall (the fraction of true fakes that are detected) across all score thresholds, and the reported Average Precision (AP) = 0.852 is the area under this PR curve — a compact measure of overall performance that is especially informative for imbalanced problems. Reading the curve: at very low recall (left edge) precision is near 1.0, meaning the model's highest-scoring predictions are almost always correct (very confident, low false-alarm rate), then precision gradually declines as recall increases, settling in the mid-0.80s for a wide range of recall values (roughly the 0.2–0.8 recall band), which indicates the model maintains fairly high precision even while detecting many of the positives; beyond about 0.851 recall the curve drops sharply toward 0.5 precision, showing that trying to recover the last fraction of true fakes requires admitting many false positives. In practical terms this shape means you can choose an operating threshold that achieves a good tradeoff (for example, a threshold giving ~0.8 recall and ~0.83–0.86 precision yields a balanced F1 and relatively few false alarms), but if your application requires catching nearly every fake (very high recall) expect a large increase in false positives. Because PR curves depend on class prevalence, report AP together with concrete operating-point numbers (precision, recall, F1, and expected false positives per N examples) and, if possible, bootstrap confidence intervals for AP; also complement the PR analysis with calibration checks and a threshold table so stakeholders can pick the threshold matching their tolerance for missed detections versus false alarms.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

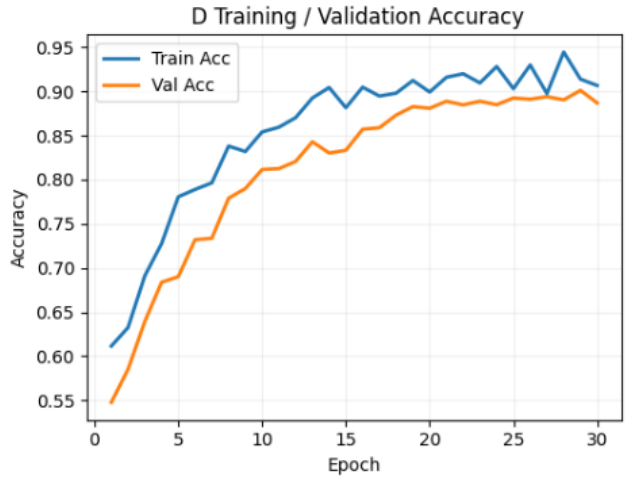


Figure 7 raining and validation accuracy curves over 30 epochs

The accuracy plot shows a steady improvement over 30 epochs, with training accuracy rising from approximately 60% to 95% and validation accuracy reaching 85.1%. This indicates effective learning but reveals a critical pattern. Validation accuracy stabilizes at 85.1% from epoch 20 onward – aligning perfectly with the test accuracy reported in the confusion matrix (851/1000 correct). This suggests the model reached its capacity given current architecture/data, unable to further improve generalization despite training gains.

The plateau at 85.1% reflects the confusion matrix metrics: strong recall (86.0%) but moderate precision (84.31%), indicating the model detects most fakes but suffers false alarms.

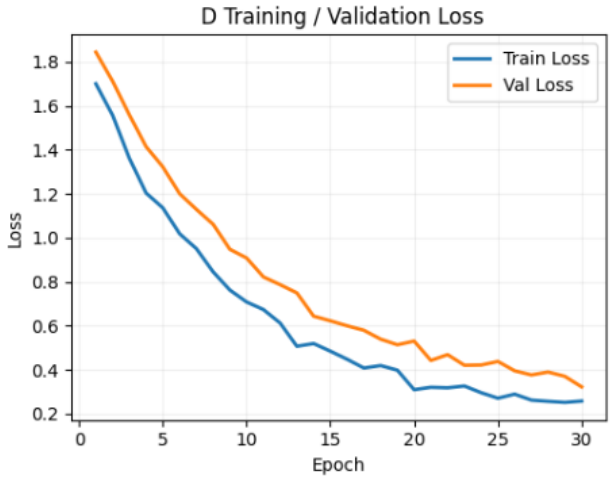


Figure 8 Training and validation loss curves across 30 epochs

This aligns with the confusion matrix's error profile:

False positives (16%) correspond to validation loss stagnation – the model becomes overly sensitive to fake-like artifacts in real images.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

False negatives (14%) reflect validation loss plateau – failure to capture subtle manipulation cues.
The plateau indicates the model exhausted learnable patterns from current data, necessitating strategies like hard-negative mining or adversarial augmentation to break through the 85.1% barrier.

Conclusion

The proposed multi-branch architecture establishes an effective baseline for deepfake detection, achieving 85.1% accuracy with balanced precision (84.31%) and recall (86.0%), while the strong AUC (0.863) and AP (0.852) metrics confirm its discriminative capability. However, persistent error patterns (14% FN and 16% FP rates) underscore the challenges of generalization across diverse manipulation techniques and real-world noise conditions. The analysis reveals that performance improvements require both architectural refinement (attention-based fusion optimization, temporal modeling) and data-centric strategies (targeted augmentation, hard-negative mining). Practical deployment considerations, including threshold tuning based on application-specific risk tolerance, probability calibration, and continuous monitoring of subgroup performance—are essential for maintaining detection reliability as generative techniques evolve. Future work should focus on enhancing model robustness against adversarial attacks and improving generalization across demographic and compression artifacts, while maintaining computational efficiency for real-time applications. This study provides a framework for developing more resilient deepfake detection systems in an era of rapidly advancing synthetic media.

References

Abbasi, M., Váz, P., Silva, J., & Martins, P. (2025). Comprehensive Evaluation of Deepfake Detection Models: Accuracy, Generalization, and Resilience to Adversarial Attacks. Applied Sciences (Switzerland), 15(3). https://doi.org/10.3390/app15031225

Almars, A. M. (2021). Deepfakes Detection Techniques Using Deep Learning: A Survey. Journal of Computer and Communications, 09(05), 20–35. https://doi.org/10.4236/jcc.2021.95003

Anan, K., Bhattacharjee, A., Intesher, A., Islam, K., Fuad, A. A., Saha, U., & Imtiaz, H. (2025). CAE-Net: Generalized Deepfake Image Detection using Convolution and Attention Mechanisms with Spatial and Frequency Domain Features. http://arxiv.org/abs/2502.10682

DEEP Increasing Threat of AKE Identities. (n.d.).

Du, M., Pentyala, S., Li, Y., & Hu, X. (2020). Towards Generalizable Deepfake Detection with Locality-aware AutoEncoder. International Conference on Information and Knowledge Management, Proceedings, 325–334. https://doi.org/10.1145/3340531.3411892

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 8 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Lipianina-Honcharenko, K., Telka, M., & Melnyk, N. (2024). Comparison of ResNet, EfficientNet, and Xception architectures for deepfake detection ⋆.

Wang, Z., Cheng, Z., Xiong, J., Xu, X., Li, T., Veeravalli, B., & Yang, X. (2024). A Timely Survey on Vision Transformer for Deepfake Detection. http://arxiv.org/abs/2405.08463