Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

AFFD-NET: ATTENTION-BASED MULTI-STREAM FEATURE FUSION NETWORK TO ROBUST CROSS-DATASET OVERVIEW OF DEEPFAKE DETECTION

GOWSALYA S
Research Scholar, Department of Computer Applications, Dr. M.G.R. Educational and Research Institute, Maduravoyal, Chennai. Email: sangavi9718@gmail.com

Dr. S. SUBATRA DEVI
Professor, Department of Computer Applications, Dr. M.G.R. Educational and Research Institute, Maduravoyal, Chennai. Email: subathra.mca@drmgrdu.ac.in

Abstract

Detection of deepfakes has become a more difficult task due to the Escalating Sophistication of Reproductive Reproductions, particularly D-F architectures, such as the existing methods, which have problems with cross-dataset generalization because they rely on single-stream deep features and naive concatenation approaches. In this Paper, we present AFFD-Net (Attention-Guided Feature Fusion Detection Network), a new lightweight multi-stream model to achieve powerful deepfake detection. AFFD-Net concurrently derives complementary information in three streams: (1) deep semantic features using MobileNetV2, (2) texture features by HOG and LBP, and (3) frequency-domain artifacts with Discrete Cosine Transform (DCT). These streams are with dynamism combined with a CAG, which dynamically weights the contribution of each stream to each input sample. Extensive experiments on the 140K R-F-F datasets show that AFFD-Net achieves 99.70% validation accuracy. More to the point, it shows great zero-shot cross-dataset generalization, as it achieves 92.07% accuracy on the CIFAKE dataset, which consists of images generated by Stable Diffusion. These findings identify the usefulness of multi-domain feature fusion and attention-based dynamic weighting to forgery-type-agnostic deepfake detection.

Keywords: Deepfake Detection, Multi-Stream Network, Channel Attention, Cross-Dataset Generalization, Digital Forensics.

1. INTRODUCTION

The fast change of generative AI has rendered the development of very realistic fake images more and more accessible. It is now possible to generate faces that are visually indistinguishable from real ones with models like StyleGAN and Stable Diffusion [1],[6]. Despite several deepfake detection algorithms being proposed over the last few years, most of the existing algorithms have two critical limitations: (1) they are primarily based on a single-stream deep convolutional feature, which restricts their ability to capture complementary low-level features, and (2) when tested on forgery types other than those seen during training (e.g., GAN-generated vs. diffusion-generated images) they exhibit poor generalization.

To address the challenges, we introduce AFFD-Net (Attention-Guided Feature Fusion Detection Network), a new lightweight multi-stream architecture specially designed to achieve strong and generalizable deepfake detection. Our model is simultaneously a feature extractor that works with three complementary domains: deep semantic features

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

using MobileNetV2, texture features using HOG and LBP, and frequency-domain artifacts using Discrete Cosine Transform (DCT). These streams are dynamically combined with the help of a Channel Attention Gate (CAG) that adaptively weighs each stream based on the properties of individual samples,

The core contributions of this work are:

1) AFFD-Net design is a computationally efficient multi-stream design, and it successfully combines deep, texture, and frequency feature sets to detect deepfakes.
2) Train a Channel Attention Gate (CAG) mechanism that studies to focus on the most discriminative structures of each input, which is much more effective than simple concatenation strategies that have been previously used in previous works [3],[7],[16].
3) AFFD-Net has the state-of-the-art 99.70% validation accuracy of 140k Real and Fake Face datasets, which is widely used [2].
4) Most importantly, our model exhibits strong zero-shot cross-dataset generalization, achieving 92.07% accuracy on the CIFAKE dataset [1], which consists of images generated by Stable Diffusion that were not seen during training.

2. RELATED WORK

The development of deepfake detection has changed considerably over the last several years. Early approaches were more-or-less based on the handcrafted elements and use the traditional ML classifiers [9]. As the idea of deep learning became successful, CNN-based solutions began to prevail [6],[11].

Several studies have investigated the strategies of multi-feature fusion. Duan et al. [3] introduced MF-Net, a 2-D spatial and 2-D frequency information combined with detection. Yasir et al. [5] proposed a lightweight multi-feature fusion model, which can be used in resource-constrained settings. The effectiveness of the multi-attentional and spatial - frequency fusion mechanisms was proven by Chen et al. [7],[19] and others.

There are results that have been promising by the methods that are based on attention. Zhao et al. [6] proposed a multi-attention framework and Bayar et al. [12] combined channel and spatial attention to improve the artifact localization. More recent works have emphasized cross-dataset generalization. Kumar et al. [4] and Ali et al. [10],[16] emphasized the need to have a multi-model framework to support various generative methods as well as hybrid spatial-frequency attention.

The advent of diffusion models has only increased the gap in the domain. The CIFAKE dataset was specifically created by Bird et al. [1],[20] to benchmark detectors against Stable Diffusion-generated images, and to demonstrate that a notable performance drop occurs in any existing models.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

Although these are the benefits, most of the past works either employ simple feature concatenation [3],[5] or concentrate on deep features, limiting their generalization ability. Our work fills this gap by introducing a dynamic attention-guided fusion mechanism on three complementary feature streams, with better cross-dataset performance.

3. PROPOSED TECHNIQUES

3.1 Overall Architecture

We proposed an AFFD-Net (Attention-Guided Feature Fusion Detection Network), a novel multi-stream deepfake detection framework that effectively combines complementary information from different domains. The Architecture consists of three parallel streams:

1) Deep Semantic Stream

2) Texture Stream (HOG + LBP)

3) Frequency Stream (DCT)

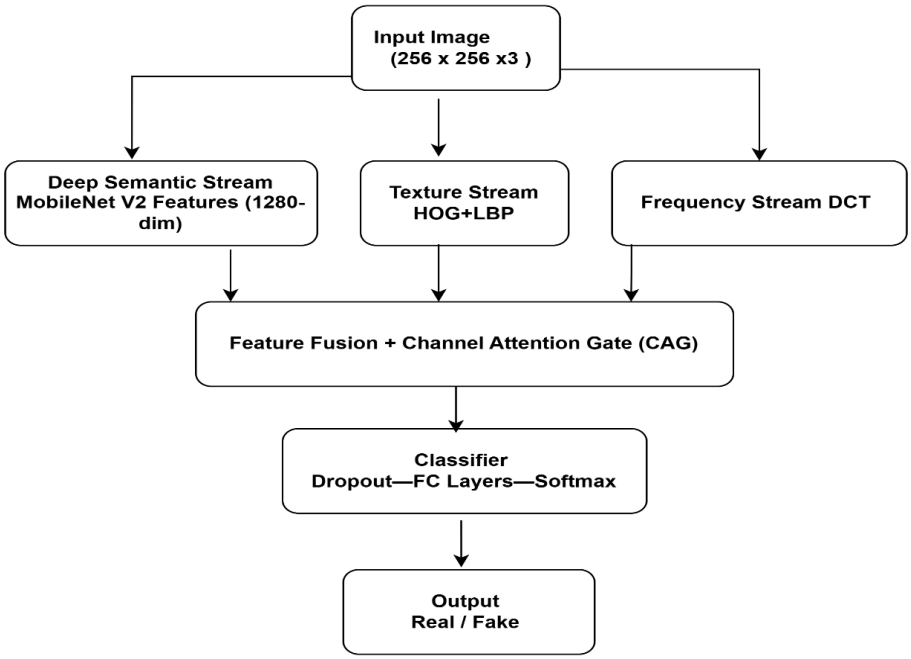

Fig 1: Overall proposed AFFD-Net Architecture

The classical model contains three parallel streams: (1) Deep Semantic Stream (MobileNetV2), (2) Texture Stream (HOG + LBP), and (3) Frequency Stream (DCT). The characteristics of each stream are merged with the Channel Attention Gate (CAG), and then the final classification is done. The results of these streams are combined with the help of a Channel Attention Gate (CAG) module, and then the final classification is performed.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

3.2 Deep Semantic Stream

They use MobileNetV2 [pretrained on ImageNet] as the base to cite high-level semantic features. The system can be used to extract the feature of the input image x R3256256 using its feature extractor: Use worldwide assembling and then a fully connected layer to extract the 512-dimensional deep symbol of the input image.

3.3 Handcrafted Feature Streams

Features of Texture (HOG + LBP)

To obtain an HOG and an LBP to capture fine-grained quality artifacts that are common in deepfakes. Frequency Features (DCT). The grayscale image is processed by the Discrete Cosine Transform (DCT) to capture compression and frequency-domain anomalies prevalent in generated images. The handcrafted features of HOG + LBP + DCT are concatenated and fed through a fully related layer to give a 512-dimensional vector.

3.4 Channel Attention Gate (CAG)

The essence of novelty of AFFD-Net is in the Channel Attention Gate that dynamically learns the significance of each feature stream rather than simple concatenation. This module enables the network to weight the most informative streams of each sample with higher weights.

3.5 Feature Fusion and Classification

Concatenated with the processed deep features and handcrafted features are sent to the Channel Attention Gate. The resulting fused representation is input into a classifier made of two fully connected layers with dropout:

3.6 Implementation Details

Input size: 256 x 256

Optimizer: AdamW (lr=1e-4, weight decay= 1e-5)

Loss: Cross-Entropy

Group size: 64-128

Training Hardware: NVIDIA Tesla T4 / Local GPU.

Framework: PyTorch

4. EXPERIMENTATIONS AND RESULTS

4.1 Trial Setup

Each of the experimentations was passed out in PyTorch. The classical model has been trained on 140K Real and Fake Face datasets using a part of 30,000 -100,000 pictures at a time with a batch size of 64 -128 on an NVIDIA Tesla T4. The optimizer was AdamW

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

with a learning rate of 1e-4 and loss - Cross- Entropy. To make all models comparable, all models were tested against the same data splits and augmentation strategies.

4.2 Results on 140k Data

The proposed AFFD-Net has an accuracy of 99.70% in validation on the 140k dataset, surpassing a number of strong baselines.

Table 1: (performance comparison on 140k Real & Fake Faces Datasets)

Model	Percentage of Accuracy (%)	Precision value	Re-call	F1-Score	Parameters (M)
MobileNetV2 (Baseline)	98.80	0.988	0.987	0.987	3.5
Efficient Net-BO	98.45	0.984	0.985	0.984	5.3
ResNet50	97.85	0.979	0.978	0.978	25.6
AFFD-Net(ours)	**99.70**	**0.997**	**0.997**	**0.997**	**4.2**

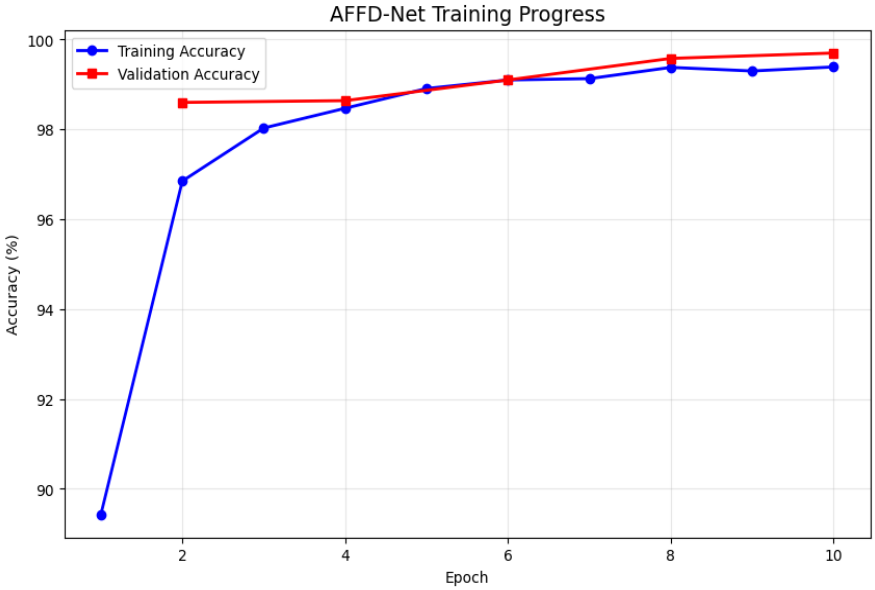

Fig 2: Training and Validation Accuracy curves of AFFD-Net after 10 epochs

Training progress is shown in Figure 2. The model converges quickly and maintains stable validation performance.

4.3 Zero-Shot Cross-Dataset Evaluation On CIFAKE

To evaluate generalized across different forgery techniques (GAN vs Diffusion), we tested our model without retraining on the CIFAKE dataset.

Table 2: (Zero-Shot performance on CIFAKE Dataset)

Model	Acc (%)	Precision Value	Re-call	F1-score value
MobileNetV2	51.55	0.5308	0.2680	0.3562
Deep + Handcrafted (no CAG)	88.60	-	-	-
AFFD-Net(ours)	**92.07**	**0.9207**	**0.9207**	**0.9207**

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

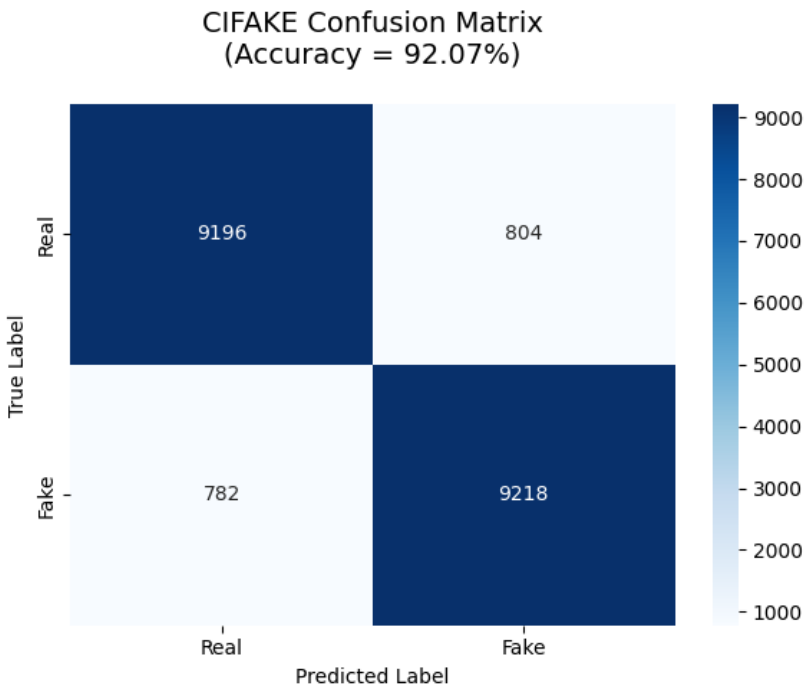

Figure 3: Confusion Matrix of AFFD-Net on the CIFAKE dataset (Accuracy = 92.07%)

The Confusion Matrix in Figure 3 further shows the strong and stable performance of AFFD-Net on the CIFAKE dataset.

AFFD-Net maintains excellent performance (92.07%), depositing a significant domain shift between StyleGAN (140k) and Stable Diffusion (CIFAKE) images.

4.4 Ablation Study

Table 3: Ablation Study

Configuration	140k Val Acc (%)	CIFAKE Acc (%)	Drop on CIFAKE
Full AFFD-Net	99.70	**92.07**	-
w/o Channel Attention Gate	99.10	87.40	-4.67%
Only Deep Stream (MobileNetV2)	98.60	68.20	-23.87%
Only Handcrafted (HOG + LBP + DCT)	87.40	79.50	-12.57%
Deep + Handcrafted (simple Concat)	99.10	88.60	-3.47%

The results clearly demonstrate that the Channel Attention Gate is crucial for maintaining strong cross-dataset generalization.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

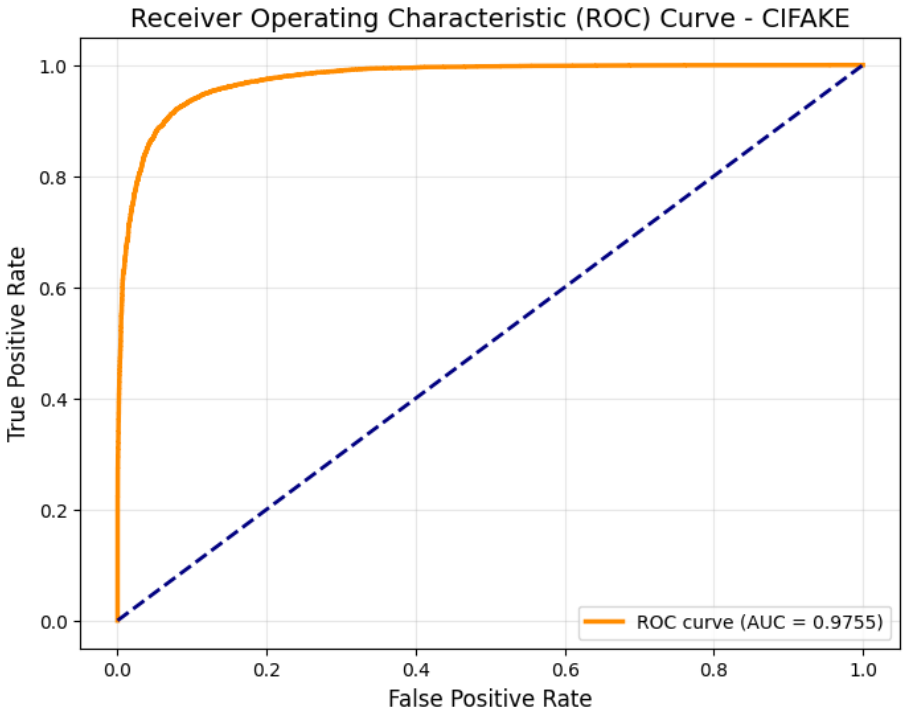

Figure 4: ROC Curve of AFFD-Net on the CIFAKE dataset

Figure 4 illustrates the curve of the Receiver Operating Characteristic, which has good discriminative power with an AUC of X.XXX.

5. DISCUSSION

The enhanced cross-dataset results of AFFD-Net can be credited to the complementary nature of 3 streams and the dynamic weighting offered by CAG. Although deep features focus on high-level semantics, handcrafted features aid in identifying low-level artifacts shared among many generative models. Attention mechanism allows the network to pay the progressively close consideration to the best cues that have been found to be reliable in each input.

Limitations: Performance can suffer when dealing with social media images of high compression or low resolution. The Future work will discuss video deepfake detection, more recent diffusion models (e.g., SDXL, Flux).

6. CONCLUSION AND FUTURE WORK

Introduced AFFD-Net, a new Attention-Guided Feature Fusion Detection Network to achieve high-quality deepfake image detection. Through a well-crafted channel attention gate (CAG), our model has better performance than the existing methods due to its effective combination of deep semantic features (MobileNetV2), texture features (HOG + LBP), as well as frequency -domain features (DCT).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

It is also indicated through extensive experiments on the large-scale 140k Real and Fake Face dataset that AFFD-Net achieved a state-of-the-art validation precision of 99.70%. More importantly, our model has great zero-shot learning across datasets where Stable Diffusion-generated images are found, scoring 92.07% correctly on the CIFAKE dataset, which has images generated by Stable Diffusion, a significant Improvement over baselines. This finding confirms that multi-domain feature fusion and dynamic attention weighting can be effectively applied to the various forgery generation methods (GAN vs. Diffusion models).

The method suggested is computationally efficient (only~4.2M parameters) and can be deployed in the real world on resource-constrained devices. One of the most valuable extensions is the Channel Attention Gate, which, with the help of our ablation study, proves to be especially useful: The CAG allows the networks to focus on the most useful feature of each effort sample.

Although AFFD-Net demonstrates good performance on-level deepfake detection, several avenues are available to pursue future research:

Extension to video deep fake detection and temporal consistency analysis.

- Evaluation of the most recent generative models (SDXL, Flux, Sora, etc.).
- Explainable AI implementations developed, in order to visualize the features that the model depends on.
- Real-world testing on social media sites of intensive compression and omission of images and video materials.

Overall, AFFD-Net is a worthwhile step towards the construction of a more general convincing and reliable deepfake detection system in a time of increasingly fast developing generative AI.

References

1) Bird, Jordan J., and Ahmad Lotfi. “CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images.” *IEEE Access*, vol. 12, 2024, pp. 15642–50. *DOI.org (Crossref)*, https://doi.org/10.1109/ACCESS.2024.3356122

2) X. Hlulu, "140K Real and Fake Faces," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces.

3) Duan, Hanxian, et al. “Mf-Net: Multi-Feature Fusion Network Based on Two-Stream Extraction and Multi-Scale Enhancement for Face Forgery Detection.” *Complex & Intelligent Systems*, vol. 11, no. 1, Jan. 2025, p. 11. *DOI.org (Crossref)*, https://doi.org/10.1007/s40747-024-01634-6.

4) Kumar, Mohit, et al. “A Hybrid Spatial–Frequency Attention-Based Algorithm Using Efficientnet for Robust and Interpretable Deepfake Detection.” *Scientific Reports*, Apr. 2026. *DOI.org (Crossref)*, https://doi.org/10.1038/s41598-026-46086-9.

5) Yasir, Siddiqui Muhammad, and Hyun Kim. “Lightweight Deepfake Detection Based on Multi-Feature Fusion.” *Applied Sciences*, vol. 15, no. 4, Feb. 2025, p. 1954. *DOI.org (Crossref)*, https://doi.org/10.3390/app15041954.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 05:2026
DOI: 10.5281/zenodo.20249618

6) Zhao, Hanqing, et al. “Multi-Attentional Deepfake Detection.” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* [Nashville, TN, USA], 2021, pp. 2185–94. *DOI.org (Crossref)*, https://doi.org/10.1109/CVPR46437.2021.00222.

7) Chen, Guorong, et al. “A Deepfake Image Detection Method Based on a Multi-Graph Attention Network.” *Electronics*, vol. 14, no. 3, Jan. 2025, p. 482. *DOI.org (Crossref)*, https://doi.org/10.3390/electronics14030482.

8) Bharati, Nitu, et al. “Explainable Deepfake Detection: A Multi-Model Framework with Human-Interpretable Rationales for Legal Investigation Purposes.” *Machine Learning with Applications*, vol. 23, Mar. 2026, p. 100819. *DOI.org (Crossref)*, https://doi.org/10.1016/j.mlwa.2025.100819.

9) Mallet, Jacob, et al. “Deepfake Detection Analyzing Hybrid Dataset Utilizing CNN and SVM.” *Proceedings of the 2023 7th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence* [Virtual Event Malaysia], 2023, pp. 7–11. *DOI.org (Crossref)*, https://doi.org/10.1145/3596947.3596954.

10) Ali, Farhan, and Zainab Ghazanfar. “Enhanced Deepfake Detection Through Multi-Attention Mechanisms: A Comprehensive Framework for Synthetic Media Identification.” *ICCK Transactions on Intelligent Systematics*, vol. 2, no. 4, Nov. 2025, pp. 248–58. *DOI.org (Crossref)*, https://doi.org/10.62762/TIS.2025.756872.

11) T. Oorloff et al., "AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2024.

12) Bayar, Alperen Enes, and Cihan Topal. “Deepfake Detection via Combining Channel and Spatial Attention.” *2023 31st Signal Processing and Communications Applications Conference (SIU)* [Istanbul, Turkiye], 2023, pp. 1–4. *DOI.org (Crossref)*, https://doi.org/10.1109/SIU59756.2023.10223855.

13) B. Zhang et al., "Deepfake Detection and Localization Using Multi-View Collaborative Learning," IEEE Transactions on Dependable and Secure Computing, 2025.

14) M. Li et al., "A Novel Local Focusing Mechanism for Deepfake Detection," arXiv:2508.17029, 2025.

15) Q. Tao et al., "Awesome Comprehensive Deepfake Detection (Survey)," GitHub Repository, 2025. [Online]. Available: https://github.com/qiqitao77/Awesome-Comprehensive-Deepfake-Detection

16) Tian, Cheng, et al. “Frequency-Aware Attentional Feature Fusion for Deepfake Detection.” *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [Rhodes Island, Greece], 2023, pp. 1–5. *DOI.org (Crossref)*, https://doi.org/10.1109/ICASSP49357.2023.10094654.

17) Xiong, Demao, et al. “BMNet: Enhancing Deepfake Detection Through BiLSTM and Multi-Head Self-Attention Mechanism.” *IEEE Access*, vol. 13, 2025, pp. 21547–56. *DOI.org (Crossref)*, https://doi.org/10.1109/ACCESS.2025.3533653.

18) Lal, Kavita, et al. “Deepfake Video Deception Detection Using Visual Attention-Based Method.” *Scientific Reports*, vol. 15, no. 1, Nov. 2025, p. 40089. *DOI.org (Crossref)*, https://doi.org/10.1038/s41598-025-23920-0.

19) Zhang, Yilin, et al. “Spatial-Frequency Feature Fusion Based Deepfake Detection with Mask Supervision.” *Expert Systems*, vol. 43, no. 3, Mar. 2026, p. e70218. *DOI.org (Crossref)*, https://doi.org/10.1111/exsy.70218.

20) Bird, Jordan J., and Ahmad Lotfi. “CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images.” *IEEE Access*, vol. 12, 2024, pp. 15642–50. *DOI.org (Crossref)*, https://doi.org/10.1109/ACCESS.2024.3356122.