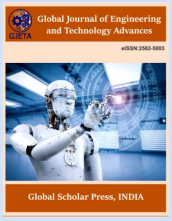
Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/



(RESEARCH ARTICLE)

Check for updates

Comparative evaluation of lightweight CNN models for deepfake detection under social media compression conditions

Olawunmi Asake Adebanjo [1,*] and Ernest E. Onuiri [2]

[1] *Department of Software Engineering, School of Computing, Babcock University, Ilishan-Remo, Ogun State, Nigeria.*
[2] *Department of computer science, School of Computing, Babcock University, Ilishan-Remo, Ogun State, Nigeria.*

Global Journal of Engineering and Technology Advances, 2026, 27(02), 042-050

Publication history: Received on 29 March 2026; revised on 05 May 2026; accepted on 08 May 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.27.2.0111

Abstract

The proliferation of deepfake media on social media platforms demands detection solutions that balance computational efficiency with robustness under realistic media conditions. While state of the art detection models such as XceptionNet and EfficientNet-B4 achieve high accuracy, their computational intensity limits real-time deployment at scale. This study presents a comparative evaluation of two lightweight convolutional neural network (CNN) architectures, MobileNetV2 (3.4M parameters) and ShuffleNetV2 (1.4M parameters), for deepfake detection under social media compression conditions. Both architectures were evaluated using 5-fold stratified cross-validation on a combined dataset of FaceForensics++ (C23 compression) and Celeb-DF v2, comprising approximately 13,529 videos and 295,976 face crops. Results demonstrate that ShuffleNetV2 outperforms MobileNetV2 across all key metrics (test AUC: 0.7839 vs 0.7674, accuracy: 86.17% vs 85.85%) while utilising 59% fewer parameters and training 23% faster. A transfer learning ablation study confirms that ImageNet pre-training is essential, as training from scratch yielded a near random performance (AUC: 0.5966). Compression robustness analysis at Q95 and Q50 JPEG quality levels reveals that both models maintain overall performance degradation below 3%, though ShuffleNetV2 exhibits a notable 5.33% AUC drop on FaceForensics++ at low quality. All models exceed real-time requirements by wide margins (>190 FPS at batch size 1 on A100 GPU). These findings establish lightweight CNN architectures as viable candidates for scalable deepfake detection and provide empirical evidence for selecting efficient architectures for social media deployment.

Keywords: Deepfake detection; Lightweight CNN; MobileNetV2; ShuffleNetV2; Social media; Compression robustness; Transfer learning; Real-time detection

1. Introduction

The digital landscape has changed rapidly in recent years, particularly with increased reliance on social media platforms where millions of people now consume and share content daily [1]. This environment has become fertile ground for deepfake propagation. Since the first deepfake surfaced on Reddit in 2017, the tools used to create and modify digital media have continually evolved [2]. With readily available techniques and internet tutorials, it is now possible to create hyper-realistic manipulated media [3]. Research indicates that false media can proliferate on social platforms six times more quickly than accurate information [4], making effective deepfake detection essential as the prevalence of deepfakes continues to grow [5].

Detecting deepfakes in real-world conditions presents substantial challenges that extend beyond controlled laboratory settings. When media is shared across social media platforms, it undergoes various forms of degradation including compression, resolution reduction, and format conversion. These transformations can suppress the subtle manipulation traces that detection systems rely upon, significantly reducing detection accuracy. Studies have shown that deepfake

* Corresponding author: Olawunmi A. Adebanjo

detection models trained on high quality benchmark datasets often experience considerable performance drops when evaluated on compressed or degraded media typical of social media environments [6]. This gap between benchmark performance and real-world applicability represents a great obstacle for practical deployment.

Convolutional Neural Networks (CNNs) have shown great promise in identifying subtle visual differences that are frequently invisible to the human eye, allowing for accurate and rapid identification of fake content [7]. CNNs, with their ability to identify specific artifacts and distortions, are well-suited to separating authentic from synthetic content. However, social media platforms house large volumes of user-generated content, requiring detection solutions that can operate efficiently at scale [8]. State of the art deepfake detection models, while achieving high accuracy on benchmark datasets, typically rely on computationally intensive architectures such as XceptionNet [9] and EfficientNet-B4 [10] that require significant processing power, limiting their applicability for large scale real-time content screening.

Lightweight CNN architectures offer potential solutions to these computational constraints, yet they remain insufficiently studied for deepfake detection under degraded media conditions. Two architectures of particular interest are MobileNetV2 [11], which employs depthwise separable convolutions and inverted residual blocks to minimise computational cost, and ShuffleNetV2 [12], which utilises channel shuffling to enable efficient cross group information exchange. Both architectures have demonstrated strong performance in general visual recognition tasks, but their comparative effectiveness for deepfake detection under social media compression conditions has not been rigorously evaluated.

This study addresses this gap through a systematic comparative evaluation of MobileNetV2 and ShuffleNetV2 for deepfake detection. The specific contributions of this paper are as follows: (1) a rigorous comparison of two lightweight CNN architectures for deepfake detection using 5-fold stratified cross-validation across two benchmark datasets; (2) a compression robustness analysis evaluating detection performance under high-quality (Q95) and low-quality (Q50) JPEG compression conditions; (3) a transfer learning ablation study demonstrating the critical role of ImageNet pre-training for lightweight deepfake detectors; and (4) a comprehensive computational efficiency analysis establishing real-time viability for social media deployment.

2. Related work

2.1. CNN-Based Deepfake Detection

Several CNN-based approaches have been proposed for deepfake detection, varying significantly in computational requirements and detection strategies. Afchar et al. [13] designed MesoNet, a compact model incorporating convolutional layers paired with max-pooling layers and inception units, achieving 95% detection rate for face-to-face videos and 98% for deepfake videos under real-world conditions. At the heavier end, Chollet [9] introduced the Xception architecture that employs depthwise separable convolutions to efficiently separate spatial and channel-wise correlations, achieving strong classification accuracy. Rössler et al. [14] evaluated multiple architectures including VGG16, ResNet, and XceptionNet on the FaceForensics++ benchmark, establishing XceptionNet as a strong baseline for deepfake detection.

More recent work has explored hybrid architectures. Jasim et al. [15] combined ResNet and DenseNet with evolutionary optimization, achieving 99.75% accuracy on a 140K image dataset. Hsu et al. [16] employed DenseNet in a Siamese network architecture with contrastive loss, achieving 98.8% precision. Temporal approaches have also been used, with Masud et al. [17] developing LW-DeepFakeNet, a lightweight time-distributed CNN-LSTM network achieving 99.24% accuracy with real-time processing capability. Khormali et al. [18] developed a Vision Transformer-based framework achieving 99.41% accuracy on FaceForensics++ but with substantially higher computational demands. While these approaches achieve high accuracy, they generally require substantial computational resources, limiting their suitability for real-time social media deployment.

2.2. Lightweight Architectures for Visual Recognition

MobileNetV2 [11] introduced inverted residual blocks with linear bottlenecks, factorising standard convolutions into depthwise and pointwise operations to reduce computational cost from $O(D^{K^2} \times M \times N \times D^{F^2})$ to $O(D^{K^2} \times M \times D^{F^2} + M \times N \times D^{F^2})$. ShuffleNetV2 [12] takes a different approach, following four practical design guidelines for efficient architectures: equal channel width to minimise memory access cost, avoiding excessive group convolutions, reducing network fragmentation, and accounting for element-wise operations. Its channel shuffling mechanism rearranges feature map channels after grouped convolutions, enabling cross-group information exchange without additional computational overhead.

Both architectures have been widely adopted for mobile and edge deployment in general computer vision tasks. However, their relative effectiveness specifically for deepfake detection, and particularly under the compression conditions characteristic of social media platforms, has not been systematically compared.

2.3. Compression Robustness in Deepfake Detection

The impact of media compression on deepfake detection performance is a recognised challenge. Studies have demonstrated that detection models trained on high-quality benchmark datasets often experience significant performance degradation when applied to compressed media [6]. Compression artefacts can mask or alter the subtle features that models use to identify deepfakes [19]. This is particularly relevant for social media environments where content undergoes multiple compression cycles during upload, processing, and delivery. Most existing lightweight approaches either sacrifice significant detection accuracy or lack rigorous evaluation across diverse compression conditions [20]. The extent to which lightweight architectures can maintain acceptable detection accuracy under such conditions remains an insufficiently explored question, motivating the compression robustness analysis presented in this work.

3. Methodology

3.1. Research Design

This study employs an experimental research design to comparatively evaluate two lightweight CNN architectures for deepfake detection. A structured pipeline consisting of model selection, training with transfer learning, and comprehensive evaluation under social media-like conditions was followed. The primary objective is to determine which lightweight architecture offers the best balance of detection accuracy, compression robustness, and computational efficiency for practical social media deployment.

3.2. Datasets

The study utilises two benchmark datasets widely adopted in deepfake detection research. The FaceForensics++ (FF++) dataset [14] contains approximately 7,000 videos with C23 compression covering four manipulation techniques: Deepfakes, Face2Face, FaceSwap, and NeuralTextures. The C23 compression setting introduces moderate compression artefacts similar to those found in user-generated social media content. The Celeb-DF v2 dataset [21] contains 6,529 videos featuring 59 celebrities, generated using an improved synthesis process that significantly reduces visual artefacts, making detection more challenging and providing a rigorous test for model generalisation. Table 1 summarises the dataset configuration.

Table 1 Dataset Summary

Dataset	Total Videos	Compression	Real Videos	Fake Videos
FaceForensics++ (C23)	7,000	Medium (C23)	1,000	6,000
Celeb-DF v2	6,529	Raw/Light	590	5,939

Both datasets were split independently using identical proportions: 70% for training, 10% for validation, and 20% for testing. Stratification ensured equal representation of real and fake videos in each split, applied per dataset to eliminate class imbalance. Video-level splitting was enforced so that all frames from a given video remain in the same split, preventing data leakage. The combined test set comprises 77,284 face crops (38,134 from FF++ and 39,150 from Celeb-DF).

3.3. Data Preprocessing

A consistent preprocessing pipeline was applied across both datasets. Frames were extracted at a rate of 30 uniformly sampled frames per video. Face detection was performed using MTCNN [22] with a minimum confidence threshold of 0.9 and a 20% margin around detected faces. Detected faces were resized to 224 × 224 pixels and normalised to pixel values in the range [0, 1]. During training, data augmentation was applied including random horizontal flips ($p = 0.5$), rotation (±15°), brightness and contrast adjustments (±20%), random cropping (90–100%), and Gaussian noise ($\sigma = 0.01$). To simulate the multiple compression cycles typical of social media sharing, additional JPEG compression was applied to a subset of training images with quality factors ranging from 70 to 95.

3.4. Model Architectures

Both architectures were implemented using PyTorch 2.0+ with official ImageNet pre-trained weights from torchvision (IMAGENET1K_V1). A unified framework ensured consistent data augmentation, preprocessing, and training procedures for fair comparison. For MobileNetV2, the first 100 layers were frozen, and the classification head was replaced with a dropout layer (rate 0.5) followed by a linear layer with a single output for binary classification. The same modification was applied to ShuffleNetV2 with the first 50 layers frozen. Both models used BCEWithLogitsLoss as the loss function. Table 2 summarises the architectural comparison.

Table 2 Baseline Architecture Comparison

Metric	MobileNetV2	ShuffleNetV2
Parameters	3.4M	1.4M
Model Size	~14 MB	~6 MB
FLOPs (224×224)	~300M	~150M
Key Innovation	Inverted Residuals	Channel Shuffle
Frozen Layers	First 100	First 50
Added Layers	Dropout(0.5) → Linear(1)	Dropout(0.5) → Linear(1)

3.5. Training Configuration

Both models were trained using 5-fold stratified cross-validation to obtain robust performance estimates with standard deviation. Hyperparameters were selected via grid search: learning rate from {0.0001, 0.001, 0.01}, batch size from {16, 32, 64}, dropout rate from {0.3, 0.5, 0.7}, and weight decay from {1e-4, 1e-5, 1e-6}. The optimal configuration used a learning rate of 0.001, batch size of 32, dropout rate of 0.5, and weight decay of 1e-5 with the Adam optimiser. A ReduceLROnPlateau scheduler monitored validation loss and reduced the learning rate by a factor of 0.1 when no improvement was observed for 5 consecutive epochs (minimum learning rate: 1e-7). Early stopping halted training if validation AUC did not improve for 10 consecutive epochs, restoring the best model weights.

3.6. Evaluation Framework

Models were evaluated using standard classification metrics: accuracy, precision, recall, F1-score, and AUC-ROC. Compression robustness was assessed by testing models at two JPEG quality levels: high quality (Q95) and low quality (Q50), measuring the percentage AUC drop between conditions. Per-dataset evaluation (FF++ and Celeb-DF separately) assessed cross-dataset generalisation. Computational efficiency was benchmarked on standardised hardware (2× NVIDIA A100-PCIE-40GB on Vast.ai) measuring total parameters, model size, inference time, and frames per second (FPS) at batch sizes of 1, 8, and 32. A transfer learning ablation study compared models initialised with ImageNet pre-trained weights against random initialisation.

4. Results and discussion

4.1. Experimental Setup

All experiments were conducted on 2× NVIDIA A100-PCIE-40GB GPUs using PyTorch 2.10.0 with Automatic Mixed Precision (AMP), with a fixed random seed of 42 for reproducibility. Training was conducted in subprocess isolation with one model per GPU in parallel. The combined dataset yielded 295,976 face crops from approximately 13,529 videos.

4.2. Baseline Cross-Validation Results

Table 3 presents the 5-fold cross-validation results for MobileNetV2. The model achieved a mean validation AUC of 0.7878 ± 0.0044 with a mean validation accuracy of 86.69%, requiring 3.26 hours for the complete 5-fold training cycle.

Table 3 MobileNetV2 5-Fold Cross-Validation Results

Fold	Val AUC	Val Accuracy	Time
1	0.7906	86.90%	0.58h
2	0.7882	86.87%	0.62h
3	0.7797	86.25%	0.56h
4	0.7880	86.60%	0.73h
5	0.7925	86.82%	0.77h
Mean	0.7878 ± 0.0044	86.69%	3.26h total

Table 4 presents the corresponding results for ShuffleNetV2, which achieved a mean validation AUC of 0.8052 ± 0.0062 and mean validation accuracy of 86.88%, completing 5-fold training in 2.50 hours.

Table 4 ShuffleNetV2 5-Fold Cross-Validation Results

Fold	Val AUC	Val Accuracy	Time
1	0.8037	86.72%	0.44h
2	0.8098	87.15%	0.51h
3	0.7936	87.46%	0.52h
4	0.8094	86.47%	0.50h
5	0.8093	86.62%	0.52h
Mean	0.8052 ± 0.0062	86.88%	2.50h total

Table 5 provides a comprehensive comparison of both models on the held-out test set. ShuffleNetV2 outperforms MobileNetV2 across all key metrics while utilising 59% fewer parameters. It achieves notably higher test AUC on Celeb-DF (0.7946 vs 0.7666) and comparable AUC on FF++ (0.7682 vs 0.7676). ShuffleNetV2 also trains 23% faster (2.50h vs 3.26h for 5-fold cross-validation). These results demonstrate that the channel shuffling mechanism in ShuffleNetV2 is more effective than MobileNetV2's inverted residuals for deepfake feature extraction.

Table 5 Baseline Model Comparison Summary

Metric	MobileNetV2	ShuffleNetV2
Mean Val AUC	0.7878 ± 0.0044	0.8052 ± 0.0062
Mean Val Accuracy	86.69%	86.88%
Test AUC (FF++)	0.7676	0.7682
Test AUC (Celeb-DF)	0.7666	0.7946
Test AUC (Overall)	0.7674	0.7839
Test Accuracy	85.85%	86.17%
Test F1	0.9201	0.9227
Parameters	3.4M	1.4M
Training Time (5-fold)	3.26h	2.50h

4.3. Transfer Learning Ablation Study

To quantify the importance of transfer learning for lightweight deepfake detection, ShuffleNetV2 was additionally trained from scratch without pre-trained weights. Table 6 presents the results. Training from scratch resulted in near-random performance (AUC: 0.5966, accuracy: 20.82%), with the model predicting all samples as a single class and

demonstrating zero discrimination ability. With ImageNet pre-trained weights, the same architecture achieves strong discriminative performance (AUC: 0.8052), representing an absolute improvement of +0.2086 in AUC. This confirms that transfer learning is essential for lightweight deepfake detection, as the pre-trained features provide a critical foundation for learning manipulation-specific patterns.

Table 6 Transfer Learning Ablation

Configuration	Val AUC	Val Accuracy	Result
From scratch	0.5966	20.82%	Failed
ImageNet pre-trained	0.8052	86.88%	Success
Improvement	+0.2086	+66.06%	—

4.4. Per-Dataset Evaluation

Table 7 presents the per-dataset test performance at high quality (Q95). Both models perform better on Celeb-DF than FF++. MobileNetV2 achieves similar AUC on both datasets (0.7676 on FF++ vs 0.7666 on Celeb-DF), while ShuffleNetV2 shows a more pronounced difference (0.7682 on FF++ vs 0.7946 on Celeb-DF). The stronger Celeb-DF performance likely reflects the more distinctive artefacts in that dataset's deepfakes, while FF++ C23 compression obscures manipulation traces. ShuffleNetV2 leads on both datasets, with its advantage being particularly pronounced on Celeb-DF (+0.0280 AUC).

Table 7 Per-Dataset Test Performance (Q95)

Model	FF++ AUC	FF++ Acc	CDF AUC	CDF Acc	Δ
MobileNetV2	0.7676	84.15%	0.7666	87.50%	-0.001
ShuffleNetV2	0.7682	84.36%	0.7946	87.94%	+0.026

4.5. Compression Robustness Analysis

Tables 8 and 9 present the compression robustness analysis. At the overall level (Table 8), both models maintain performance degradation below 3% when compression quality drops from Q95 to Q50. MobileNetV2 shows a 1.37% AUC drop while ShuffleNetV2 degrades by 2.91%. The per-dataset analysis (Table 9) reveals a more nuanced picture. ShuffleNetV2 exhibits a notable 5.33% AUC drop on FF++ at low quality, substantially larger than MobileNetV2's 1.52% drop on the same dataset. In contrast, ShuffleNetV2 is more robust on Celeb-DF, with only a 0.87% drop compared to MobileNetV2's 1.40%. This suggests that ShuffleNetV2's channel shuffling mechanism may be more sensitive to compression artefacts that interact with FF++'s existing C23 compression, while MobileNetV2's depthwise separable convolutions provide more uniform robustness across compression conditions.

Table 8 Performance Across Compression Levels (Overall)

Model	High (Q95)	Low (Q50)	Δ AUC	Δ %
MobileNetV2	0.7674	0.7569	-0.0105	-1.37%
ShuffleNetV2	0.7839	0.7611	-0.0228	-2.91%

Table 9 Per-Dataset Compression Analysis

Model	Dataset	High (Q95)	Low (Q50)	Δ %
MobileNetV2	FF++	0.7676	0.7559	-1.52%
MobileNetV2	Celeb-DF	0.7666	0.7559	-1.40%
ShuffleNetV2	FF++	0.7682	0.7272	-5.33%
ShuffleNetV2	Celeb-DF	0.7946	0.7877	-0.87%

4.6. Computational Efficiency Analysis

Table 10 presents the computational efficiency comparison. MobileNetV2 achieves slightly faster single-image inference (3.75ms vs 5.14ms) despite having 2.4× more parameters, likely due to more optimised depthwise separable convolution GPU kernels. However, both models exceed real-time requirements (>30 FPS) by wide margins at all batch sizes. At batch size 1, MobileNetV2 achieves 267 FPS while ShuffleNetV2 achieves 194 FPS. At batch size 32, the gap narrows substantially (6,211 vs 5,978 FPS), indicating that ShuffleNetV2's efficiency advantages become more pronounced in high-throughput batch processing scenarios typical of social media content pipelines.

Table 10 Computational Efficiency Comparison (A100 GPU)

Metric	MobileNetV2	ShuffleNetV2
Parameters	3.4M	1.4M
Inference (batch=1)	3.75ms	5.14ms
FPS (batch=1)	267	194
FPS (batch=8)	2,169	1,519
FPS (batch=32)	6,211	5,978
Training Time (5-fold)	3.26h	2.50h

5. Discussion

The results of this comparative evaluation yield several findings with implications for the design and deployment of lightweight deepfake detection systems.

First, ShuffleNetV2's shows consistent superiority over MobileNetV2 across detection metrics, despite having 59% fewer parameters, this suggests that the channel shuffling mechanism is more effective than inverted residuals for learning deepfake discriminative features. The channel shuffle operation enables richer cross group information flow, which may be particularly beneficial for detecting the subtle inter-channel inconsistencies introduced by deepfake generation processes. This finding challenges the assumption that larger parameter counts necessarily translate to better detection performance within the lightweight architecture family.

Secondly, the transfer learning ablation study provides strong evidence that ImageNet pre-training is not merely beneficial but essential for lightweight deepfake detection. The near-random performance of ShuffleNetV2 trained from scratch (AUC: 0.5966) indicates that the model's limited capacity prevents it from learning meaningful representations from deepfake data alone. The pre-trained features provide a critical foundation of low-level and mid-level visual representations upon which manipulation-specific patterns can be built. This has practical implications: deploying lightweight deepfake detectors requires access to pre-trained weights, and the quality of these initial representations directly constrains the achievable detection performance.

Finally, the compression robustness analysis reveals an important vulnerability. While both models maintain overall degradation below 3%, ShuffleNetV2's 5.33% AUC drop on FF++ at Q50 represents a practically significant weakness. This dataset-specific vulnerability suggests that ShuffleNetV2's learned features may include compression-sensitive patterns that are disrupted by additional JPEG compression layered on top of FF++'s existing C23 compression. MobileNetV2's more uniform robustness across datasets and compression levels (1.37–1.52% drops) indicates that its depthwise separable convolutions may learn more compression invariant representations. This finding highlights the importance of evaluating detection models not just at a single compression level but across a range of conditions representative of social media environments.

The following limitations are acknowledged. The AUC values in the 0.76–0.79 range, while demonstrating meaningful discrimination ability, fall below the performance of heavyweight models such as XceptionNet [9] and Vision Transformers [18] that achieve AUC values above 0.95 on similar benchmarks. This represents the fundamental trade-off between computational efficiency and detection accuracy. Additionally, this study evaluates frame-level detection without temporal modelling, this may limit detection of manipulation artifacts that manifest across frames. The evaluation is also limited to two benchmark datasets; the performance on wild deepfakes may differ.

Future work will investigate knowledge distillation techniques to address the compression robustness gap identified in ShuffleNetV2, exploring whether transferring learned representations from MobileNetV2 can improve ShuffleNetV2's resilience to compression artefacts while preserving its efficiency advantages. Additionally, temporal modelling approaches and evaluation on more diverse, in-the-wild datasets will be pursued.

6. Conclusion

This study presented a systematic comparative evaluation of two lightweight CNN architectures, MobileNetV2 (3.4M parameters) and ShuffleNetV2 (1.4M parameters), for deepfake detection under social media compression conditions. Using 5-fold stratified cross-validation across FaceForensics++ and Celeb-DF v2 datasets comprising approximately 295,976 face crops, the evaluation provides robust evidence for architectural selection in resource-constrained deployment scenarios.

ShuffleNetV2 emerged as the superior architecture, outperforming MobileNetV2 on all key metrics (test AUC: 0.7839 vs 0.7674, accuracy: 86.17% vs 85.85%) while utilising 59% fewer parameters and training 23% faster. The transfer learning ablation study confirmed that ImageNet pre-training is essential for lightweight deepfake detection, with from-scratch training yielding near-random performance. Both models exceed real-time requirements by wide margins, establishing their viability for social media deployment.

However, ShuffleNetV2's compression vulnerability on FF++ at low quality (5.33% AUC drop) identifies a specific weakness that warrants further investigation. Addressing this robustness gap through techniques such as knowledge distillation represents a promising direction for developing lightweight detection models that are both accurate and resilient under the diverse compression conditions encountered on social media platforms.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] R. Mubarak, T. Alsboui, O. Alshaikh, I. Inuwa-Dutse, S. Khan, and S. Parkinson, "A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats," IEEE Access, vol. 11, pp. 144497–144529, 2023.

[2] T. T. Nguyen, C. M. Nguyen, D. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection," ArXiv, abs/1909.11573, 2019.

[3] A. Ismail, M. Elpeltagy, M. S. Zaki, and K. Eldahshan, "A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost," Sensors, vol. 21, no. 16, 2021.

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[5] M. Taeb and H. Chi, "Comparison of Deepfake Detection Techniques through Deep Learning," Journal of Cybersecurity and Privacy, vol. 2, no. 1, pp. 89–106, 2022.

[6] H. Ajder, G. Patrini, F. Cavalli, and L. Cullen, "The State of Deepfakes: Landscape, Threats, and Impact," Deeptrace, 2019.

[7] A. Tiwari, R. Dave, and M. Vanamala, "Leveraging Deep Learning Approaches for Deepfake Detection: A Review," in ACM International Conference Proceeding Series, pp. 12–19, 2023.

[8] M. S. Rana, M. Solaiman, C. Gudla, and M. F. Sohan, "Deepfakes – Reality Under Threat?," in 2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC), pp. 721–727, 2024.

[9] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258, 2017.

[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510–4520, 2018.

[12] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in European Conference on Computer Vision (ECCV), pp. 116–131, 2018.

[13] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–7, 2018.

[14] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1–11, 2019.

[15] H. K. Jasim, L. A. Hassnawi, and H. A. Hashem, "Deepfake Image Detection Using ResNet and DenseNet with Evolutionary Optimisation," in AIP Conference Proceedings, 2024.

[16] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep Fake Image Detection Based on Pairwise Learning," Applied Sciences, vol. 10, no. 1, p. 370, 2020.

[17] M. Masud, G. H. Eldin, M. Alshammari, A. Alabdulkarim, and N. Alghamdi, "LW-DeepFakeNet: A Lightweight Time Distributed CNN-LSTM Network for Real-Time Deepfake Detection," Applied Soft Computing, vol. 151, p. 111137, 2024.

[18] A. Khormali and J.-S. Yuan, "DFDT: An End-to-End DeepFake Detection Framework Using Vision Transformers," Applied Sciences, vol. 12, no. 6, p. 2953, 2022.

[19] M. Masood, M. Nawaz, K. Malik, A. Javed, A. Irtaza, and H. Malik, "Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward," Applied Intelligence, vol. 53, pp. 1–53, 2022.

[20] A. H. Soudy et al., "Deepfake detection using convolutional vision transformers and convolutional neural networks," Neural Computing and Applications, vol. 36, no. 31, pp. 19759–19775, 2024.

[21] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3204–3213, 2020.

[22] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.