International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 07, Issue 03, August 2026
e-ISSN: 2745-9659
https://ijcis.net/index.php/ijcis/index

Analysis of the Effect of Social Media Compression on the Accuracy of Deepfake Detection Using MobileNetV3 and Compressed Data Augmentation

1st Erika Ramadhani, 2nd M. Fahrul Ramadhan
1st *Department of Informatics*, 2nd *Magister of Informatics*
Universitas Islam Indonesia
Yogyakarta, Indonesia
erika@uii.ac.id

***Abstract*— *Deepfake content is increasingly difficult to distinguish from authentic media and is widely distributed through social media platforms, where compression can remove visual traces used by deepfake detectors. However, the robustness of lightweight deepfake detection architectures, particularly MobileNetV3-Small, under varying compression levels remains poorly characterized. This study investigates MobileNetV3-Small degradation under controlled compression conditions and evaluates compressed-data augmentation as a low-cost mitigation strategy. We conducted a quantitative experiment using the Celeb-DF v2 dataset under three compression conditions: raw, c23, and c4, simulated with FFmpeg. We trained MobileNetV3-Small on raw data and evaluated it across all three compression levels, then compared it with a mixed-training scenario combining raw, c23, and c40 data. Evaluation on 5,180 face images from the official test partition showed that accuracy decreased from 70.23% on raw data to 69.54% at c23 and 58.11% at c40. The sharpest degradation occurred between c23 and c40, mainly driven by a drop in recall from 89.79% to 61.24%, while precision remained relatively stable. Compressed-data augmentation improved AUC across all compression levels and increased accuracy at c40 from 58.11% to 61.72%, although it introduced a slight accuracy trade-off on raw data. These findings empirically characterize how compression affects a lightweight deepfake detector and show that incorporating compressed samples during training can improve robustness under heavy compression. The results have practical implications for developing lightweight deepfake detection systems intended for resource-constrained and compression-affected media environments.*

Keywords*: Deepfake detection, MobileNetV3, Image Compression, Data Augmentation, Celeb-DF.*

I. INTRODUCTION

The development of generative deep learning techniques, especially generative adversarial networks (GANs) and autoencoders, has enabled the creation of deepfake content that is increasingly realistic and accessible to the public [1], [2]. The misuse of deepfakes for disinformation, defamation, and identity fraud makes deepfake detection an important digital security issue for individuals, media organizations, and law enforcement agencies [1]. In real-world distribution, deepfake content is often accessed after social media platforms such as Instagram, WhatsApp, Facebook, and Twitter/X process it, applying platform-specific compression to reduce bandwidth and storage requirements [3].

Such compression can remove fine-grained visual artifacts and frequency patterns that pixel-based deepfake detection models commonly exploit. Consequently, models that perform well under laboratory conditions may degrade substantially when evaluated on compressed content [3]. This difference between performance under controlled laboratory conditions and real-world distribution conditions represents an important generalization challenge in deepfake detection [4]. The problem is particularly relevant for detection systems intended to operate in resource-constrained environments, including on-device content moderation, mobile applications, and mobile camera-based identity verification. These applications require detection models that balance detection performance and computational

efficiency, motivating the use of lightweight convolutional neural network (CNN) architectures such as MobileNetV3 [5].

This study selected MobileNetV3 because it was designed to balance accuracy and latency efficiently for mobile devices [5]. Its lightweight design combines depthwise separable convolution, squeeze-and-excitation mechanisms, and hard-swish activation to reduce computational requirements while maintaining effective feature extraction [5]. These characteristics make MobileNetV3 a relevant architecture for investigating whether a computationally efficient deepfake detector can maintain its detection capability when the input quality is degraded by compression. Moreover, previous lightweight deepfake detection studies have shown that MobileNetV3 can achieve high detection performance with relatively low memory and computational requirements [6]. However, these studies have primarily evaluated the architecture under high-quality input conditions and have not systematically characterized its behavior across different compression levels.

Several approaches have been proposed to address compression-related degradation. Montibeller et al. [3] developed a framework that emulates a sequence of video-sharing processes on social media and demonstrated that retraining with emulated data can improve performance under real-world distribution conditions. Lagsoun et al. [7] proposed GA-LASSO feature selection to maintain detection performance on compressed content, while Lai et al. [8] investigated contrastive learning across compression levels. Other studies have explored frequency-domain representations, which have been reported to improve robustness to domain shifts and compression compared with purely pixel-based approaches [9]. Spatial-frequency collaborative architectures [10] and lightweight frequency-domain detectors for low-resolution imagery [11] further show the potential of frequency-based information to improve robustness.

For lightweight deepfake detection, Harshitha et al. [6] proposed a MobileNetV3-based detector enhanced with a lightweight attention mechanism and reported an F-score of up to 98.46% with low memory and computational requirements. Kumar et al. [12] demonstrated the effectiveness of combining residual connections, squeeze-and-excitation, and depthwise convolution for efficient deepfake detection, while Xia et al. [13] improved a lightweight MesoNet detector through a preprocessing module. Nevertheless, these studies do not specifically address how a lightweight detector behaves when the same test content is subjected to progressively stronger compression.

The research gap addressed in this study is therefore twofold. First, although researchers have investigated compression robustness in deepfake detection, systematic evaluation has largely focused on larger architectures, while the degradation behavior of MobileNetV3-Small across controlled compression levels remains insufficiently characterized [3], [4]. Second, previous lightweight deepfake detection studies have generally emphasized detection performance under high-quality conditions rather than examining the relationship between compression severity, detection degradation, and the underlying error pattern [6]. This creates a need to quantify not only how much performance decreases, but also which detection component compression affects most.

Accordingly, this study makes three specific contributions. First, it empirically characterizes MobileNetV3-Small performance degradation across raw, c23, and c40 compression conditions using the Celeb-DF v2 dataset. Second, it analyzes the degradation using accuracy, precision, recall, F1-score, AUC, and confusion matrices to identify the dominant error pattern as compression severity increases. Third, it evaluates compressed-data augmentation as a low-cost mitigation strategy by comparing a raw-trained model with a model trained using combined raw and compressed data. Through these experiments, the study empirically characterizes the compression robustness of a lightweight deepfake detector and provides practical evidence for improving its reliability in compression-affected media environments.

II. RESEARCH METHODS

2.1 Research Stages

This research consisted of seven successive stages: dataset collection, frame extraction and face cropping, compression simulation, video-level data partitioning, MobileNetV3-Small

training, evaluation, and result analysis. Figure 1 presents the overall research workflow.

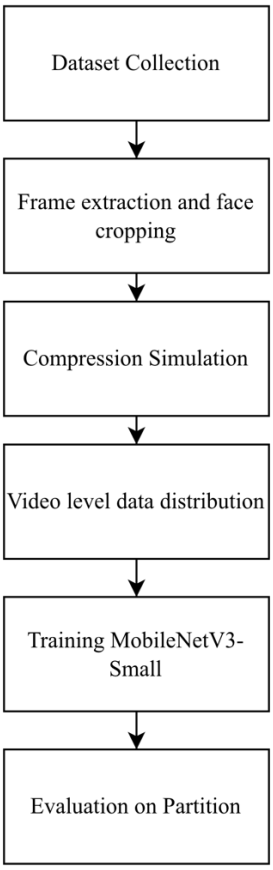

Figure 1. Research Method

The first stage involved collecting the Celeb-DF v2 dataset [14]. The dataset contains 590 Celeb-real videos, 300 YouTube-real videos, and 5,639 Celeb-synthesis deepfake videos, together with an official test list containing 518 videos. We selected the dataset because Celeb-DF v2 provides relatively realistic deepfake content and is widely used for deepfake detection evaluation. For the training and validation subsets, we sampled videos at the video level to maintain class balance, using 135 videos per class for training and 15 videos per class for validation. We retained the official Celeb-DF v2 test partition for final evaluation.

The second stage involved extracting temporal frames and cropping faces. We uniformly sampled ten frames from each video across its temporal duration. We used uniform sampling to ensure the extracted frames represented different temporal portions of each video rather than concentrating samples in a particular segment. We used a fixed sampling configuration of 10 frames per video to provide consistent representation across videos while keeping the number of processed images computationally manageable. The extracted frames were subsequently processed using Multi-task Cascaded Convolutional Networks (MTCNN) to detect and crop the facial region. The resulting face crops constituted the image-level units used in subsequent experiments.

Before using them in the detection model, the extracted face images were resized to 224 × 224 pixels to match the MobileNetV3-Small input dimension. We normalized the images using the ImageNet normalization configuration associated with the pretrained MobileNetV3-Small model. We kept the MTCNN configuration and detection parameters fixed across all experimental conditions to ensure that differences in detection performance were attributable to the compression and training scenarios rather than changes in preprocessing.

The third stage simulated controlled compression conditions using FFmpeg. We evaluated three conditions: raw, representing the original data without additional compression; c23, representing H.264 compression with a Constant Rate Factor (CRF) of 23; and c40, representing H.264 compression with a CRF of 40 combined with a 0.75× downscaling operation. The c23 and c40 labels follow the quality-level convention commonly used in FaceForensics++ [15]. We treated these settings as controlled compression conditions rather than exact reproductions of proprietary social-media processing pipelines.

The fourth stage partitioned the data at the video level rather than the frame level. This procedure prevented frames from the same source video from appearing in different partitions, reducing the risk of data leakage. The training-to-validation ratio was 90:10, while the test partition followed the official Celeb-DF v2 test list. Table 1 presents the resulting data distribution.

Table 1. Distribution of Data

Split	Video per class	Facial Image Real	Facial Image Fake
Train	135	1.350	1.350
Validation	15	150	150
Test	178 Real / 340 Fake	1.780	3.400

The fifth stage trained MobileNetV3-Small under two scenarios: train-raw and train-mix. The

train-raw scenario used only raw training images, whereas train-mix combined raw, c23, and c40 training images. We initialized MobileNetV3-Small with ImageNet-pretrained weights via torchvision. We froze the backbone parameters and retrained only the classifier head. The input size was fixed at 224 × 224 pixels. We optimized the model with Adam (learning rate 1×10^{-4}, batch size 32) for up to 10 epochs. Early stopping with a patience of 3 epochs was applied to prevent unnecessary training after validation performance stopped improving.

The sixth stage evaluated both trained models on the official test partition under all three compression conditions, resulting in six model–compression combinations. Performance was measured using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC), together with confusion matrices to characterize the types of classification errors.

The seventh stage involved analyzing the performance changes across compression conditions, identifying the observed region of sharp performance degradation, examining the contribution of false positives and false negatives through the confusion matrices, comparing the train-raw and train-mix scenarios, and deriving practical implications and limitations.

2.2 MobileNetV3

MobileNetV3 is a family of lightweight *convolutional neural network* architectures designed by Howard et al. [5] through a combination of *platform-aware neural architecture search* and manual enhancements to achieve a balance between accuracy and latency on *mobile devices*. This architecture's efficiency comes from three main components: *a residual inverted* block with *depthwise separable convolution*, a *squeeze-and-excitation* module that adaptively rebalances each feature channel's contribution, and a *hard-swish activation function* [5]. In the deepfake detection domain, Harshitha et al. [6] achieved an F-score of 98.46% with MobileNetV3 enriched with light attention; Kumar et al. [12] demonstrated the effectiveness of combining *residual connections* with *squeeze-and-excitation;* and Xia et al. [13] amplified the light detector through a pre-processing module.

2.3 Compressed Data Augmentation

Compressed data augmentation is a training strategy that adds a copy of the compressed training data to the training set so that the model learns deepfake markers that survive the loss of pixel detail. The c23 and c40 compression level conventions follow the FaceForensics++ quality partition [15]. The effectiveness of a similar data-driven approach has been demonstrated by Montibeller et al. [3] through *fine-tuning* on social media emulation data, while Lai et al. [8] take the *contrastive learning* path across compression levels. Unlike the two, this study implemented augmentation by combining raw, c23, and c40 data into a single set of trainers without changing the architecture or loss functions, resulting in low adoption costs.

III. RESULT AND ANALYSIS

3.1 Experimental Evaluation Results

The evaluation was conducted using two training scenarios and three compression conditions. The first scenario, referred to as train-raw, was trained exclusively on raw training data, whereas the second scenario, referred to as train-mix, was trained using a combination of raw, c23, and c40 data. Both models were evaluated on the same official Celeb-DF v2 test partition containing 5,180 facial images under the three compression conditions, resulting in six model–compression combinations.

We evaluated performance using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). We also analyzed confusion matrices to determine whether performance degradation was primarily due to false positives or false negatives.

The evaluation addressed two questions. First, how does increasing compression severity affect the detection performance of a lightweight MobileNetV3-Small detector when the model is trained only on high-quality data? Second, can exposure to compressed samples during training reduce the resulting performance degradation? We address the first question through the train-raw results in Sections 3.2 and 3.3, and examine the second by comparing train-raw and train-mix in Section 3.5.

3.2 Raw-Training Scenario under Compression

Table 2 presents the evaluation results for the MobileNetV3-Small model trained exclusively on raw data. The results show that compression affected performance differently across the three conditions. Accuracy decreased only slightly from 70.23% under raw conditions to 69.54% at c23, representing a reduction of 0.69 percentage points. However, accuracy decreased substantially from 69.54% at c23 to 58.11% at c40, corresponding to an additional reduction of 11.43 percentage points. Overall, the model's accuracy decreased by 12.12 percentage points from raw to c40.

Table 2. Evaluation of MobileNetV3 in Raw Training Scenario

Compression	Accuration	Accuracy	Recall	F1 Score	AUC
Raw/high quality	70,23	71,87	89,79	79,84	69,28
Medium compression (c23)	69,54	71,60	88,82	79,29	68,06
High compression (c40)	58,11	70,96	61,24	65,74	59,51

This result indicates that the detector remained relatively stable under the c23 condition but became substantially less reliable under the heavier c40 compression. Therefore, the degradation cannot be interpreted simply as a gradual loss of performance as compression increases. Instead, the results suggest a transition to substantially poorer detection performance under the heaviest compression condition.

Decomposing the evaluation metrics provides a clearer explanation of this degradation. Precision remained relatively stable across the three conditions, decreasing only from 71.87% at raw to 70.96% at c40. In contrast, recall decreased from 89.79% at raw to 88.82% at c23 and then sharply to 61.24% at c40. The 27.58-percentage-point drop in recall from raw to c40 indicates that heavy compression primarily reduced the model's ability to identify existing deepfake samples.

This distinction matters for deepfake detection because lower recall means more manipulated samples are classified as authentic. In a detection or forensic screening context, such false negatives matter because compressed deepfake content may pass through the detector without being flagged. The relatively stable precision indicates that the degradation was not primarily caused by a growing tendency to label authentic content as fake.

The F1-score followed recall, decreasing from 79.84% at raw to 79.29% at c23 and 65.74% at c40. AUC also decreased from 69.28% at raw to 68.06% at c23 and 59.51% at c40. The reduction in AUC indicates that heavy compression affected not only the final classification decision but also the model's ability to discriminate between fake and real samples based on its output scores.

Taken together, these results indicate that heavy compression substantially weakens the forensic evidence available to the pixel-based lightweight detector. The key observation is therefore not simply that accuracy decreases, but that the degradation is dominated by a collapse in deepfake recall while precision remains comparatively stable.

3.3 Confusion Matrix and Error Analysis

Table 3 provides further evidence of how compression changes the train-raw model's error profile. Under raw conditions, the model correctly identified 3,053 of the 3,400 deepfake images, resulting in 347 false negatives. At c23, the number of false negatives increased only slightly to 380, indicating that the model retained a relatively similar ability to detect deepfake samples under medium compression.

Table 3. Raw training scenario confusion matrix

Compression Rate	TP	FN	FP	TN
Raw/high quality	3.053	347	1.195	585
Medium compression (c23)	3.020	380	1.198	582
High compression (c40)	2.082	1.318	852	928

A substantially different pattern emerged at c40. The number of true positives decreased from 3,020 at c23 to 2,082, while false negatives increased from 380 to 1,318. Thus, the number of undetected deepfake images increased by 938 images between c23 and c40. More than one-third of the 3,400 deepfake test images were therefore classified as non-fake under the c40 condition.

Interestingly, false positives decreased from 1,198 at c23 to 852 at c40. This simultaneous increase in false negatives and decrease in false positives indicates a systematic shift in the model's predictions toward the real class under heavy compression. In other words, the model did not simply become less accurate; its decision behavior changed as compression severity increased.

This shift helps explain the sharp decline in recall observed in Table 2. Heavy compression appears to reduce the visual information that the model relies on to distinguish manipulated from authentic facial content. As the discriminative cues become less available, the model increasingly assigns compressed deepfake samples to the real class. This interpretation is consistent with the observed increase in false negatives and the corresponding decrease in true positives.

The error distribution also highlights an important distinction between detection performance and forensic risk. Although the number of false positives decreases at c40, this apparent improvement in specificity-related behavior does not indicate better detection performance. Instead, it occurs simultaneously with a large increase in false negatives. For a deepfake screening system, this behavior is potentially more problematic because manipulated content can remain undetected.

The relatively high number of false positives under raw and c23 conditions should also be interpreted in the context of the test distribution, which contains 3,400 fake and 1,780 real images, as well as the limited learning capacity associated with the frozen backbone. Therefore, the confusion matrix suggests that both class distribution and model capacity contribute to the observed error pattern, while the sharp increase in false negatives at c40 is specifically associated with severe compression.

3.4 Training Dynamics and Overfitting Indications

The training process in both scenarios indicates limitations in learning capacity that should be noted. In the *train-raw* scenario, training stops early after the fourth epoch: *training* accuracy increases from 63.26% to 78.19%, while validation accuracy moves only from 58.33% to 61.00%, and the best validation loss occurs in the first epoch (0.6656). In the *mixed-training* scenario, training stopped after the fifth epoch with a similar pattern: *training* accuracy reached 79.60%, while validation accuracy stayed in the 58% range. This gap indicates *early overfitting* caused by the relatively small training dataset (1,350 images per class due to subsampling) and *backbone*-freezing strategies that limit model capacity. The accuracy reported in Table 2 and Table 4 is likely to improve if the training data size is increased or *thorough fine-tuning* is applied.

3.5 Effectiveness of Compressed Data Augmentation

Table 4 shows an informative *trade-off* between the two training scenarios.

Table 4. Comparison of raw and mixed training scenarios

Compression Rate	Accuracy (Raw Training)	Accuracy (Mixed Training)	AUC (Raw Train)	AUC (Mixed Train)
Raw/high quality	70,23	67,97	69,28	71,84
Medium compression (c23)	69,54	68,73	68,06	71,62
High compression (c40)	58,11	61,72	59,51	64,29

In the raw and c23 conditions, the *mixed-train* model's accuracy was slightly lower (67.97% and 68.73%) than the *raw-trained* model (70.23% and 69.54%), but in the c40 condition, it was 3.61 percentage points higher (61.72% compared to 58.11%). More importantly, the *blended trainer* model's AUC was superior at all compression levels: 71.84% compared to 69.28% at raw, 71.62% compared to 68.06% at c23, and 64.29% compared to 59.51% at c40.

Figure 2 further visualizes the accuracy and AUC trends across compression levels for both scenarios as a line graph.

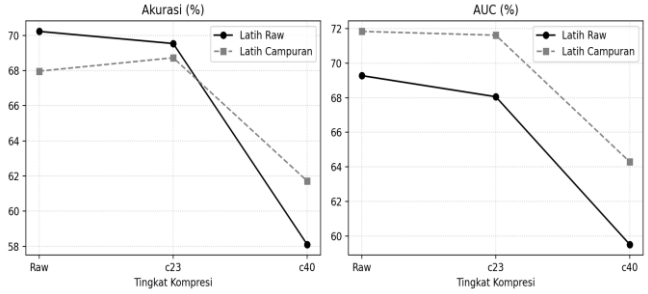

Figure 2. Accuracy and AUC trends to compression levels in both training scenarios

Figure 2 shows two main patterns. First, the accuracy curve of the two scenarios is relatively flat from raw to c23 and then drops sharply to c40, confirming the location of the critical threshold; However, the downward slope of the *mixed-train* scenario was steeper, from 68.73% to 61.72% (down 7.01 points) compared to the raw *training scenario*, which decreased by 11.43 points. Second, the AUC curve in *the mixed-train* scenario was consistently higher than the *raw-trained scenario* across the entire compression range, showing that augmentation benefits are not limited to heavily compressed conditions.

Cost-wise, augmentation scenarios only require duplicating the training data with two compressed variants generated once upfront using FFmpeg, with no architectural changes, no parameter additions, and no additional inference costs at deployment. These characteristics distinguish them from architecture-based solutions such as frequency branching [10] or *cross-compression contrastive learning* [8] which demand more complex model modifications and training processes.

3.6 Practical Implications and Research Limitations

The practical implications of this study for developers of on-device deepfake detection systems can be summarized as follows: compressed data augmentation should be the standard practice for lightweight detector training because it is low cost and has proven benefits in real distribution conditions; the critical threshold between c23 and c40 can be used as a reference for moderation policies; and the stability *of precision* in the midst of declining *recall* suggests that negative labels on heavily compressed content need to be treated with skepticism.

The limitations of this study include the use of a small subsample of trained data and *backbone* freezing that leads to *premature overfitting*, the use of a single architecture and a single dataset that limits the generalization of claims, as well as FFmpeg-based compression simulations that are not completely identical to the platform's *proprietary* process set, as emphasized by Montibeller et al. [3]. Further research should focus on thorough *fine-tuning* with larger datasets, validation using real WhatsApp and Instagram uploads, and integrating frequency-domain features into the branch [9], [10].

3.7 Discussion

This section positions the findings in the literature by comparing the results with previous research. Harshitha et al. [6]used the MobileNetV3 enriched light attention mechanism and reported an F-score of 98.46%, but their tests were limited to high-quality data without compression. Montibeller et al. [3] emulated a series of social media processes and found that *fine-tuning* the emulated data matched performance on the original upload. Lagsoun et al. [7] relied on selecting GA-LASSO features that maintained detection performance in high-compression video, while Tan et al. [9] showed that frequency-domain learning improves generalization across datasets compared to a pure pixel approach. This study tested MobileNetV3-Small with compressed data augmentation on Celeb-DF v2 raw, c23, and c40 levels, and found a critical degradation threshold between c23 and c40 marked by a decline in *recall* from 89.79% to 61.24% and showed that augmentation increased AUC at c40 from 59.51% to 64.29%.

Three main points warrant discussion. First, the accuracy and F1-score of this study were nominally lower than the results of Harshitha et al. [6], who reported an F-score of 98.46%. This difference can be explained by methodological factors: this study only trained *classifier heads* without *thorough fine-tuning* and without additional attention mechanisms, using a small subsample of training data, and was evaluated on a specially designed Celeb-DF v2 that was more challenging [14][16]. Second, the direction of the findings of this study is consistent with Montibeller et al. [3]: training on data that resembles real distribution conditions improves performance under these conditions; An additional contribution of this research is to show that similar benefits can be obtained with simple FFmpeg-based compression augmentation without having to fully emulate the platform's process suite. Third, the decline in recall in c40 aligns with the frequency-domain literature [8], [9], [16], which explains that heavy compression removes the high-frequency generative artifacts that underpin pixel-based models.

This study's findings also align with conclusions from large-scale generalization studies [4] and systematic reviews [1], [2] that detector performance is highly sensitive to shifts in test conditions relative to the training domain. This study adds to the literature by quantifying, for lightweight architectures, the degradation rate of 12.12 percentage points from raw to c40 and decomposing it into *a recall* component; this is one of the first reports for MobileNetV3-Small on the official Celeb-DF v2 Test protocol.

VI. CONCLUSION

This study demonstrates that controlled compression conditions substantially affect the performance of the lightweight MobileNetV3-Small deepfake detector. The raw-trained model showed relatively stable performance from raw to c23, with accuracy decreasing by only 0.69 percentage points, but experienced a substantially larger decrease of 11.43 percentage points from c23 to c40. This degradation was mainly driven by a sharp drop in recall from 89.79% to 61.24%, while precision remained relatively stable at about 71%. The confusion matrix further showed that false negatives increased from 380 at c23 to 1,318 at c40, indicating that heavy compression substantially increased the number of deepfake samples classified as real.

The study also demonstrates that compressed-data augmentation can improve robustness under heavy compression without modifying the model architecture. Training with combined raw, c23, and c40 data increased c40 accuracy from 58.11% to 61.72% and AUC from 59.51% to 64.29%. The mixed-training model also achieved higher AUC across all three compression conditions, though this improvement came with a small reduction in accuracy under raw conditions.

Overall, this study empirically characterizes MobileNetV3-Small performance degradation across controlled compression levels, identifies recall and false negatives as the dominant components of degradation under heavy compression, and evaluates compressed-data augmentation as a low-cost mitigation strategy. These findings indicate that training with compressed samples can improve the robustness of lightweight deepfake detection systems when input quality is degraded.

Future work should investigate these findings using larger training datasets and more extensive fine-tuning, validate the results using content processed by actual social media platforms, and examine the integration of frequency-domain features to further improve detection under extreme compression conditions.

THANK-YOU NOTE

The author expresses gratitude to the Informatics Study Program, Faculty of Industrial Technology, Islamic University of Indonesia, as well as the Digital Forensics Study Center UII for their support of facilities and scientific discussions during this research.

REFERENCES

[1] R. Ramanaharan, D. B. Guruge, and J. I. Agbinya, “DeepFake video detection: Insights into model generalisation — A Systematic review,” *Data Inf. Manag.*, vol. 9, no. 4, p. 100099, Dec. 2025, doi: 10.1016/j.dim.2025.100099.

[2] A. K. Singh and R. K. Banyal, “A Comprehensive Review of DL-Based Methods for DeepFake Image, Audio, and Video Detection,” *Int. J. Image Graph.*, Mar. 2026, doi: 10.1142/S0219467827501002.

[3] A. Montibeller, D. Shullani, D. Baracchi, A. Piva, and G. Boato, “Bridging the Gap: A Framework for Real-World Video Deepfake Detection via Social Network Compression Emulation,” in *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*, New York, NY, USA: ACM, Oct. 2025, pp. 29–36. doi: 10.1145/3746265.3759670.

[4] S. A. Khan and D.-T. Dang-Nguyen, “Deepfake Detection: Analyzing Model Generalization Across Architectures, Datasets, and Pre-Training Paradigms,”

IEEE Access, vol. 12, pp. 1880–1908, 2024, doi: 10.1109/ACCESS.2023.3348450.

[5] A. Howard *et al.*, "Searching for MobileNetV3," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.

[6] T. Harshitha, T. Simhadri, T. Jaswanthi, N. Rayvanth, and R. P. Singh, "Deepfake Image Detection Using Light-Weight Attention Integrated MobileNetV3 Model," 2025, pp. 130–142. doi: 10.1007/978-3-031-84543-7_11.

[7] A. M. Lagsoun *et al.*, "Compressed deepfake detection via GA-LASSO selection of deep features and machine learning models," *Sci. Rep.*, vol. 16, no. 1, p. 19708, Mar. 2026, doi: 10.1038/s41598-025-34733-6.

[8] C.-Y. Lai *et al.*, "UMCL: Unimodal-generated Multimodal Contrastive Learning for Cross-compression-rate Deepfake Detection," *Int. J. Comput. Vis.*, vol. 134, no. 1, p. 40, Jan. 2026, doi: 10.1007/s11263-025-02606-0.

[9] C. Tan, Y. Zhao, S. Wei, G. Gu, P. Liu, and Y. Wei, "Frequency-Aware Deepfake Detection: Improving Generalizability through Frequency Space Domain Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 5, pp. 5052–5060, Mar. 2024, doi: 10.1609/aaai.v38i5.28310.

[10] M. Qiao, R. Tian, Y. Zhai, and C. Li, "SFCL: A Spatial-Frequency Collaborative Learning Framework for Generalizable Deepfake Detection," in *2025 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, Sep. 2025, pp. 1–10. doi: 10.1109/IJCB65343.2025.11410680.

[11] D. Nayak, B. S. Prasad Mishra, and S. Champati Rai, "Lightweight and Efficient Deepfake Detection Using Transfer-Learned MobileNetV2," in *2025 2nd International Conference on Circuits, Power and Intelligent Systems (CCPIS)*, IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/CCPIS65231.2025.11234268.

[12] A. Kumar, "HybridNet: Advancing Deepfake Detection Through Residual, SE, and Depthwise Convolutions," *IEEE Access*, vol. 13, pp. 207541–207552, 2025, doi: 10.1109/ACCESS.2025.3640106.

[13] Z. Xia, T. Qiao, M. Xu, X. Wu, L. Han, and Y. Chen, "Deepfake Video Detection Based on MesoNet with Preprocessing Module," *Symmetry (Basel).*, vol. 14, no. 5, p. 939, May 2022, doi: 10.3390/sym14050939.

[14] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," Mar. 2020.

[15] A. Rossler, D. Cozzolino, L. Verdoliva, and ..., "Faceforensics++: Learning to detect manipulated facial images," *Proceedings of the …*, 2019, [Online]. Available: http://openaccess.thecvf.com/content_ICCV_2019/html/Rossler_FaceForensics_Learning_to_Detect_Manipulated_Facial_Images_ICCV_2019_paper.html

[16] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF (v2): a new dataset for DeepFake Forensics," *arXiv preprint arXiv:1909.12962*, 2019, [Online]. Available: https://www.researchgate.net/profile/Yuezun-Li/publication/336147158_Celeb-DF_A_New_Dataset_for_DeepFake_Forensics/links/5e1bec72a6fdcc28376e4548/Celeb-DF-A-New-Dataset-for-DeepFake-Forensics.pdf