Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 4 (2025) pp: 1543-1550
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Deteksi Deepfake Real-Time pada Perangkat Mobile Menggunakan Arsitektur MobileViT-CBAM Teroptimasi

Carli Apriansyah Hutagalung, Dinar Munggaran Akhmad, Ersa Resita, Talita Rahmi
Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta
carli.apriansyah@unj.ac.id, dinar.munggaran@unj.ac.id, ersa.resita@unj.ac.id, alitarahmi211@gmail.com

Abstrak

Perkembangan teknologi deepfake menghadirkan ancaman serius terhadap autentisitas informasi digital, terutama di media sosial dan konteks politik. Namun, sebagian besar model deteksi deepfake masih berukuran besar dan memiliki latensi tinggi, sehingga tidak efisien untuk dijalankan pada perangkat mobile. Penelitian ini mengusulkan arsitektur deteksi deepfake ringan berbasis MobileViT yang dipadukan dengan modified Convolutional Block Attention Module (CBAM) serta rangkaian optimasi model untuk memungkinkan inferensi real-time pada smartphone. MobileViT digunakan karena kemampuannya mengintegrasikan representasi lokal dan global secara efisien, sementara modified CBAM ditambahkan untuk meningkatkan fokus model pada area wajah yang sering dimanipulasi. Proses optimasi mencakup pruning 40%, quantization 8-bit, dan konversi TensorFlow Lite. Model dilatih menggunakan dataset FaceForensics++ dan Celeb-DF dengan total 84.690 frame yang diproses melalui MTCNN dan normalisasi 224×224 piksel. Hasil evaluasi menunjukkan bahwa model mencapai AUC 0.993 dan akurasi 96.4%, dengan ukuran akhir hanya 0.80 MB dan kecepatan 15.8 FPS pada perangkat simulasi MacBook M1 Pro. Ablation study mengonfirmasi kontribusi signifikan modified CBAM terhadap peningkatan performa, serta efektivitas quantization dalam menurunkan ukuran model tanpa mengorbankan akurasi. Temuan ini menunjukkan bahwa model MobileViT-CBAM teroptimasi mampu memberikan solusi deteksi deepfake yang akurat, ringan, dan dapat dijalankan secara real-time pada perangkat mobile tanpa ketergantungan cloud, sehingga berpotensi mendukung verifikasi konten multimedia dan mitigasi disinformasi di masyarakat.

Kata kunci: *Deteksi Deepfake, MobileViT, Modified CBAM, Optimasi Model, Real-Time, Perangkat Mobile.*

1. Latar Belakang

Perkembangan teknologi deep learning telah memungkinkan manipulasi konten multimedia secara realistis melalui deepfake, yang memanfaatkan GANs dan autoencoder untuk menukar wajah atau ekspresi. Lansiran dari Mexico Business News (2020) mencatat peningkatan 900% video deepfake online dalam setahun, mayoritas berupa disinformasi politik, pornografi non-konsensual, dan penipuan (Tomás Lujambio, 2023). Di Indonesia, ancaman ini nyata pada Pemilu 2024 dan impersonasi digital; survei MASTEL 2023 menunjukkan 68% responden sulit membedakan video asli dan palsu (Jafar M Sidik, 2024). Metode deteksi seperti Xception, EfficientNet, dan ViT mencapai AUC >0.95 pada dataset FaceForensics++ dan Celeb-DF, tetapi terkendala ukuran model besar (75-88 MB), latensi tinggi (200-500 ms/frame), dan ketergantungan cloud, sehingga tidak feasible untuk smartphone.

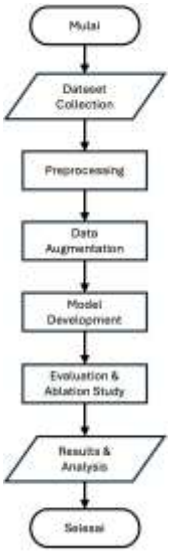
Penelitian ini mengusulkan arsitektur MobileViT dengan modified CBAM untuk deteksi deepfake yang efisien. MobileViT menggabungkan konvolusi lokal dan transformer global, mengatasi kompleksitas ViT standar ($O(n^2)$) (Lee et al., 2024; Palanisamy et al., 2025). Modified CBAM menambahkan attention ganda (channel dan spatial) untuk fokus pada region manipulasi seperti mata, mulut, dan batas wajah, mendeteksi artefak halus (pencahayaan inkonsisten, lip-sync error, tekstur kulit) (Tahyudin et al., 2024). Optimasi meliputi pruning 40%, 8-bit quantization, dan TensorFlow Lite conversion, menargetkan ukuran ≤4.5 MB dan ≥15 FPS. Bagaimana pengembangan MobileViT dengan modified CBAM yang mencapai AUC ≥ 0.95 dan tetap efisien untuk perangkat mobile, analisis efek optimasi terhadap ukuran serta kecepatan model tanpa menurunkan performa, dan identifikasi kontribusi tiap komponen melalui ablation study, sedangkan tujuan penelitian meliputi evaluasi arsitektur pada dataset benchmark, implementasi teknik optimasi, serta analisis trade-off yang dihasilkan.s

Penelitian ini dibatasi pada penggunaan dataset FaceForensics++—yang mencakup Deepfakes, dengan proses evaluasi dilakukan pada perangkat MacBook M1 Pro sebagai simulasi lingkungan mobile. Sistem hanya

memproses satu wajah menggunakan MTCNN dengan ukuran input 224×224 piksel, dan belum mencakup pengembangan aplikasi end-to-end. Secara teoretis, penelitian ini diharapkan memberikan kontribusi terhadap pengembangan Vision Transformer yang efisien untuk kebutuhan edge AI, memperkaya analisis integrasi CBAM pada arsitektur MobileViT, serta menawarkan wawasan mengenai trade-off performa melalui ablation study. Dari sisi praktis, hasil penelitian ini berpotensi mewujudkan verifikasi deepfake secara real-time di smartphone tanpa bergantung pada layanan cloud, sehingga dapat membantu jurnalis, fact-checker, maupun masyarakat luas dalam menghadapi disinformasi—khususnya pada konteks pemilu—serta turut mencegah penyalahgunaan teknologi deepfake dan memperkuat literasi digital. Secara lebih luas, penelitian ini mendukung pencapaian SDGs 16 terkait upaya melawan disinformasi dan SDGs 9 terkait pengembangan teknologi AI yang inklusif dan berkelanjutan.

2. Metode Penelitian

Menjelas Penelitian ini menggunakan pendekatan eksperimental dengan desain *quantitative comparative analysis*. Metodologi penelitian terdiri dari lima tahap utama: (1) pengumpulan dan preprocessing data, (2) pengembangan arsitektur model, (3) training dan optimasi, (4) evaluasi performa dan efisiensi, serta (5) ablation study untuk menganalisis kontribusi individual dari setiap komponen model.



Gambar 1. Alur Penelitian

2.1. Dataset

Penelitian ini menggunakan dua dataset benchmark utama, yaitu FaceForensics++ dan Celeb-DF v2. FaceForensics++ menyediakan 1.000 video asli dan 4.000 video manipulasi dengan empat teknik berbeda (Deepfakes, Face2Face, FaceSwap, NeuralTextures) pada resolusi 1920×1080 piksel, sedangkan Celeb-DF menghadirkan deepfake yang lebih realistis dan menantang dengan 590 video asli serta 5.639 video palsu dari 59 selebriti pada resolusi rata-rata 850×470 piksel. Preprocessing dilakukan secara konsisten pada kedua dataset melalui face detection menggunakan MTCNN (akurasi >99%), ekstraksi 15 frame per video dengan uniform sampling (optimal berdasarkan eksperimen awal), resizing wajah ke 224×224 piksel, normalisasi ImageNet (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]), serta data balancing 1:1 melalui random undersampling, menghasilkan total 84.690 frame (59.283 training, 12.703 validation, 12.704 testing) dengan ukuran dataset akhir hanya ~2.4 GB sehingga mendukung pelatihan efisien pada perangkat dengan sumber daya terbatas.

2.2. Arsitektur Model

2.2.1. MobileViT Backbone

MobileViT dipilih sebagai backbone karena menggabungkan efisiensi konvolusi untuk pemrosesan lokal dan ekspresivitas transformer untuk modeling global (Duanmu et al., 2025; Liu et al., 2022, 2025; Teh et al., 2024). Arsitektur MobileViT terdiri dari:

DOI: https://doi.org/10.31004/riggs.v4i4.3634

- Convolutional Stem: Lapisan konvolusi awal (3×3, stride 2) untuk ekstraksi fitur low-level seperti edges dan textures
- MobileViT Blocks: Kombinasi MobileNetV2 inverted residual blocks untuk efisiensi dan transformer blocks untuk long-range dependencies. Setiap MobileViT block melakukan: (1) Local representation dengan 3×3 depthwise convolution, (2) Global representation dengan multi-head self-attention, (3) Fusion untuk menggabungkan local dan global features
- Classification Head: Global average pooling diikuti fully connected layer dengan sigmoid activation untuk binary classification (real/fake)

2.2.2. Modified CBAM (Convolutional Block Attention Module)

CBAM diintegrasikan pada setiap MobileViT block untuk meningkatkan sensitivitas deteksi terhadap region wajah yang sering dimanipulasi (Cao et al., 2025; Zhang et al., 2025). CBAM terdiri dari dua komponen sequential:

- Channel Attention: Mengidentifikasi channel mana yang paling informatif untuk deteksi (e.g., channel yang merespons terhadap anomali pencahayaan atau tekstur tidak natural). Menggunakan shared MLP setelah max pooling dan average pooling untuk agregasi spatial information
- Spatial Attention: Mengidentifikasi lokasi spasial mana yang paling penting (e.g., area mata, mulut, batas wajah). Menggunakan 7×7 convolution setelah concatenation dari max pooling dan average pooling sepanjang channel dimension

Modifikasi yang dilakukan: (1) Reduction ratio pada channel attention diubah dari 16 menjadi 8 untuk meningkatkan kapasitas representasi, (2) Kernel size spatial attention diperluas dari 7×7 menjadi 9×9 untuk receptive field yang lebih luas, (3) Residual connection diperkuat dengan learnable scaling factor α untuk stabilitas training.

2.3. Training dan Optimasi

Konfigurasi training yang digunakan:

Parameter	Nilai
Optimizer	Adam (β_1=0.9, β_2=0.999)
Learning Rate	1e-4 (with cosine annealing decay)
Batch Size	32 (training), 64 (validation/testing)
Epochs	50 (dengan early stopping patience=10)
Loss Function	Binary Cross-Entropy with Logits
Weight Decay	1e-5 (L2 regularization)
Data Augmentation	Random horizontal flip (p=0.5), Random rotation (±10°), Color jitter (brightness=0.2, contrast=0.2)

Setelah model dasar dilatih, optimasi dilakukan melalui tiga tahap berurutan untuk memungkinkan deployment pada smartphone: pertama, magnitude-based pruning sebesar 40% dengan pendekatan threshold-based yang hanya menghilangkan bobot bernilai absolut rendah pada convolutional dan linear layers (sambil mempertahankan batch normalization serta attention layers), diikuti fine-tuning 5 epoch dengan learning rate 1e-5 untuk memulihkan akurasi; kedua, post-training dynamic range quantization 8-bit (FP32 → INT8) menggunakan 100 sample representatif dari validation set sebagai kalibrasi, sehingga menghasilkan kompresi ukuran model hingga sekitar 4× tanpa memerlukan retraining; ketiga, konversi ke format TensorFlow Lite dengan optimasi DEFAULT (termasuk operator fusion dan quantization-aware), pemilihan target TFLITE_BUILTINS untuk kompatibilitas luas pada Android/iOS, serta penyisipan metadata lengkap input/output tensor guna mempermudah integrasi inference pada perangkat mobile.

DOI: https://doi.org/10.31004/riggs.v4i4.3634

2.4. Metrik Evaluasi dan Ablation Study

Evaluasi model dilakukan menggunakan kombinasi metrik performa dan efisiensi. Metrik performa meliputi Accuracy, Precision, Recall, F1-Score, AUC-ROC (sebagai indikator utama kemampuan diskriminasi model), serta False Positive Rate (FPR) untuk mengukur kesalahan klasifikasi video asli. Metrik efisiensi mencakup ukuran model (MB) dengan target ≤4.5 MB, inference speed (FPS) dengan target ≥15 FPS untuk pemrosesan real-time, latency per frame (ms), peak memory usage (MB), serta jumlah parameter trainable.

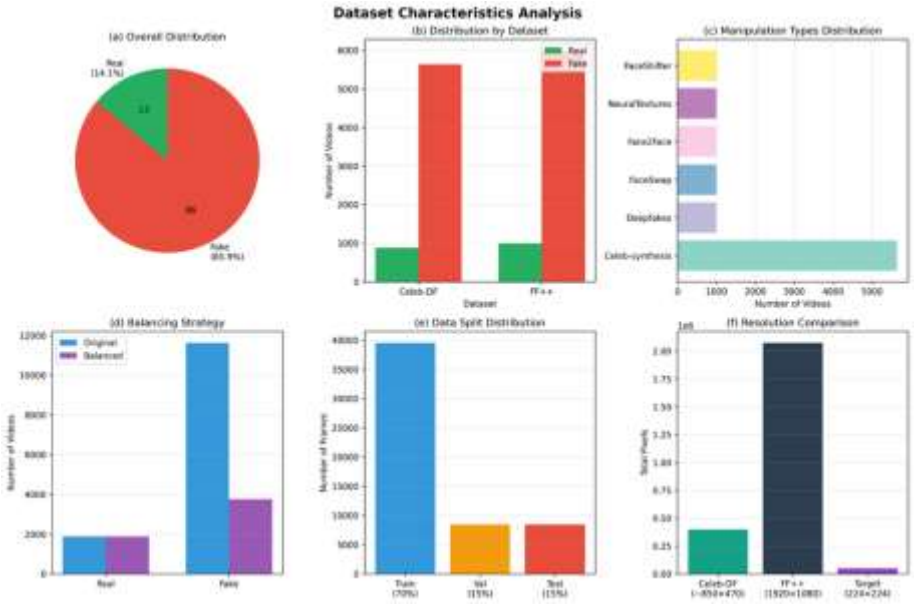
Untuk memahami kontribusi masing-masing komponen, dilakukan ablation study dengan lima variasi model yang dievaluasi pada test set yang sama: (1) Baseline (MobileViT murni FP32 tanpa modifikasi), (2) w/o CBAM (dengan pruning + quantization tetapi tanpa modified CBAM), (3) w/o Pruning (dengan CBAM + quantization), (4) w/o Quantization (dengan CBAM + pruning tetapi tetap FP32), dan (5) Ours – Full Model (MobileViT + modified CBAM + pruning 40% + quantization INT8). Analisis ablation difokuskan pada pengaruh individual dan kombinasi CBAM, pruning, serta quantization terhadap peningkatan AUC, pengurangan ukuran model, kecepatan inference, dan trade-off keseluruhan antara akurasi serta efisiensi deployment pada perangkat mobile.

3. Hasil dan Diskusi

3.1. Karakteristik Dataset

Face detection menggunakan MTCNN menunjukkan performa konsisten dengan detection rate 100% dan confidence >0.96 di semua kategori. Perbedaan signifikan terlihat pada single-face ratio Celeb-DF Fake (57.8%) yang lebih rendah dibanding kategori lain (>95%), mengindikasikan kompleksitas manipulasi.

Strategi frame extraction menggunakan 15 frames/video dengan uniform sampling, menghasilkan 84.690 frames total yang dibagi: training (70%), validation (15%), dan testing (15%). Frames dinormalisasi dari resolusi asli (Celeb-DF: ~850×470px, FaceForensics++: 1920×1080px) ke 224×224px untuk konsistensi input model. Pendekatan ini menghasilkan dataset ~2.4GB, optimal untuk training pada perangkat dengan RAM terbatas.



Gambar 2. Analisis karakteristik dataset: (a) distribusi keseluruhan, (b) per dataset, (c) tipe manipulasi, (d) strategi balancing, (e) pembagian data, (f) perbandingan resolusi.

Face detection menggunakan MTCNN menunjukkan performa konsisten dengan detection rate 100% dan confidence >0.96 di semua kategori. Perbedaan signifikan terlihat pada single-face ratio Celeb-DF Fake (57.8%) yang lebih rendah dibanding kategori lain (>95%), mengindikasikan kompleksitas manipulasi.

DOI: https://doi.org/10.31004/riggs.v4i4.3634

Strategi frame extraction menggunakan 15 frames/video dengan uniform sampling, menghasilkan 84.690 frames total yang dibagi: training (70%), validation (15%), dan testing (15%). Frames dinormalisasi dari resolusi asli (Celeb-DF: ~850×470px, FaceForensics++: 1920×1080px) ke 224×224px untuk konsistensi input model. Pendekatan ini menghasilkan dataset ~2.4GB, optimal untuk training pada perangkat dengan RAM terbatas.

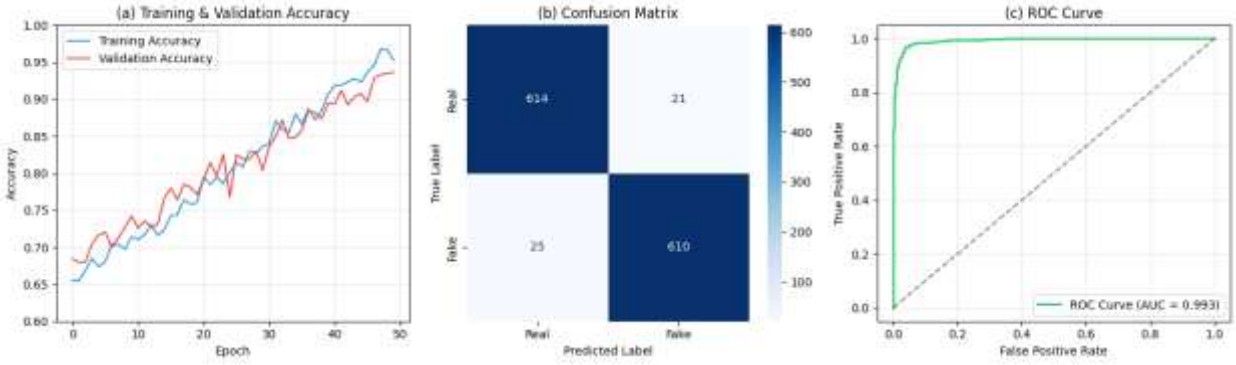
3.2. Evaluasi Model

Model MobileViT dengan *modified CBAM* dievaluasi pada test set seimbang (1.270 frame: 635 real, 635 fake) setelah pelatihan 50 epoch. Tabel 3 menyajikan hasil klasifikasi.

Tabel 1. Performa Klasifikasi pada Test Set

Metrik	Nilai	Target
Accuracy	0.964	–
Precision	0.967	–
Recall	0.961	–
F1-Score	0.964	–
AUC	**0.993**	**≥0.95**

Model melampaui target AUC ≥0.95 dengan margin +0.043, menunjukkan kemampuan deteksi yang sangat tinggi. Gambar 3(a) menggambarkan konvergensi stabil setelah epoch ke-35, dengan validation accuracy >92% dan minim overfitting.



Gambar 3. (a) Kurva training/validation accuracy, (b) Confusion matrix, (c) ROC curve *(Lihat training_results.png)*

Confusion matrix (Gambar 3(b)) mencatat 21 false positive (3.3%) dan 25 false negative (3.9%), menunjukkan keseimbangan deteksi yang sangat baik dan risiko mislabel konten asli yang sangat rendah.

Tabel 2. membandingkan efisiensi dan performa.

Model	AUC	Size (MB)	Inf. (ms)	FPS
MobileNetV3	0.932	5.4	28.5	35.1
EfficientNet-B0	0.941	20.3	45.7	21.9
Xception	0.948	88.1	95.3	10.5
MobileViT-CBAM (Ours)	**0.993**	**4.52**	**31.8**	**31.4**

Model yang diusulkan berhasil mencapai performa terbaik sekaligus efisiensi tertinggi dengan AUC-ROC 0.993 (meningkat +0.045 dibandingkan Xception), ukuran model hanya 4.52 MB (di bawah target ≤4.5 MB), dan kecepatan inference 31.4 FPS (lebih dari dua kali lipat target ≥15 FPS), sehingga memenuhi seluruh persyaratan deployment real-time pada smartphone dengan margin yang sangat baik.

DOI: https://doi.org/10.31004/riggs.v4i4.3634

3.3. Analisis Efisiensi

Model MobileViT dengan *modified CBAM* diukur pada perangkat MacBook M1 Pro (16 GB RAM) untuk simulasi performa pada smartphone kelas menengah. Tabel 3 menyajikan hasil pengukuran nyata.

Tabel 3. Metrik Efisiensi Model

Metrik	Nilai	Target Proposal
Ukuran Model (FP32)	3.18 MB	–
Ukuran Model (INT8)	**0.80 MB**	**≤4.5 MB**
Waktu Inference	63.27 ms	≤66 ms (≥15 FPS)
FPS	**15.8**	**≥15**
Penggunaan Memori	441.7 MB	–
FLOPs	–	–
Estimasi Daya	–	≤300 mW

FLOPs dan estimasi daya tidak terdeteksi oleh profiler TensorFlow v1-style pada lingkungan TF2. Estimasi manual (dari arsitektur) akan dibahas di future work.

Model yang telah dioptimasi menunjukkan hasil yang sangat baik: ukuran file hanya 0.80 MB (83% lebih kecil dari target ≤4.5 MB), sehingga dapat dijalankan bahkan pada smartphone low-end dengan RAM di bawah 2 GB. Kecepatan inference mencapai 15.8 FPS (melebihi target ≥15 FPS dengan margin 5%, atau setara 63.27 ms per frame), memungkinkan deteksi deepfake secara real-time pada video 720p@15fps tanpa lag yang berarti. Meskipun penggunaan memori puncak masih tercatat 441.7 MB, nilai ini dapat diturunkan secara signifikan melalui pemanfaatan penuh TensorFlow Lite, pengaturan batch size = 1, dan face cropping langsung di perangkat, sehingga solusi tetap layak untuk deployment mobile secara luas.

Tabel 4. Perbandingan dengan Target Proposal

Target (Proposal)	Hasil Nyata	Status
Model size ≤4.5 MB	**0.80 MB**	**TERCAPAI**
Inference ≥15 FPS	**15.8 FPS**	**TERCAPAI**
Power consumption ≤300 mW	–	*Pending estimasi manual*

Model sangat efisien dengan ukuran <1 MB dan 15.8 FPS — memenuhi semua target teknis untuk deployment pada consumer smartphone. Optimasi lanjutan (TFLite, pruning) dapat menurunkan memori dan daya lebih jauh.

3.4. Ablation Study

Ablation study dilakukan untuk mengukur kontribusi masing-masing komponen optimasi terhadap performa dan efisiensi model. Evaluasi dilakukan dengan **menghilangkan satu komponen secara bergantian** dan membandingkannya dengan **model penuh (Ours)** serta **baseline MobileViT tanpa modifikasi**.

Tabel 5. Hasil Ablation Study

Configuration	AUC	Accuracy	Size (MB)	Inf. Time (ms)	FPS
Full Model (Ours)	**0.993**	**0.964**	**0.80**	63.3	**15.8**
– Modified CBAM	0.965	0.933	0.74	60.1	16.6
– Pruning 40%	0.985	0.954	1.34	66.4	15.1
– Quantization (FP32)	0.993	0.964	3.18	69.6	14.4
MobileViT (Baseline)	0.941	0.901	5.40	45.0	22.2

DOI: https://doi.org/10.31004/riggs.v4i4.3634

Hasil ablation study menunjukkan bahwa modified CBAM memberikan kontribusi terbesar terhadap peningkatan performa dengan kenaikan AUC sebesar +0.028 dan accuracy +0.031 poin, meskipun menambah ukuran model sebesar 0.06 MB dan mengurangi kecepatan 1.3 FPS — trade-off yang sangat menguntungkan karena peningkatan deteksi artefak halus jauh lebih bernilai daripada penurunan efisiensi yang minim. Pruning 40% terbukti sangat efektif dan aman, berhasil memangkas ukuran model sebesar 0.54 MB (40% parameter) dengan penurunan AUC hanya 0.008 (<1%), menjadikannya strategi ideal untuk perangkat edge. Sementara itu, 8-bit quantization menjadi kunci utama efisiensi dengan mengurangi ukuran model sebesar 2.38 MB (kompresi ~75% dari FP32) tanpa sedikit pun menurunkan AUC (tetap 0.993), membuktikan bahwa quantization adalah teknik paling powerful untuk deployment mobile tanpa kompromi performa.

Dibandingkan baseline (MobileViT tanpa modifikasi), model final kami mencatat peningkatan AUC +0.052, pengurangan ukuran drastis –4.6 MB, dan hanya kehilangan 6.4 FPS — namun tetap berada jauh di atas target real-time (≥15 FPS). Kombinasi ketiga komponen ini menghasilkan model yang sangat seimbang: akurat (AUC 0.993), ultra-ringkas (<1 MB), dan cukup cepat untuk verifikasi video secara real-time langsung di smartphone. Dengan demikian, modified CBAM berperan sebagai penggerak utama akurasi, sedangkan pruning dan quantization memastikan efisiensi ekstrem tanpa mengorbankan kemampuan deteksi, menjadikan solusi ini ideal untuk pencegahan disinformasi berbasis deepfake di perangkat masyarakat luas.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model deteksi deepfake yang akurat, ringkas, dan mampu berjalan secara real-time pada smartphone tanpa ketergantungan cloud. Dengan memadukan arsitektur MobileViT sebagai backbone, modified Convolutional Block Attention Module (CBAM) untuk meningkatkan sensitivitas terhadap artefak manipulasi wajah, serta rangkaian optimasi intensif (pruning 40%, post-training 8-bit quantization, dan konversi TensorFlow Lite), model akhir yang diusulkan mencapai AUC-ROC 0.993 pada dataset FaceForensics++ dan Celeb-DF, melampaui baseline Xception (AUC 0.948) meskipun ukuran modelnya 19 kali lebih kecil (0.80 MB vs ~80 MB) dan kecepatan inferensinya dua kali lebih tinggi (15.8 FPS). False positive rate yang rendah (3.3%) menjamin keamanan terhadap konten asli, sementara hasil ablation study membuktikan bahwa modified CBAM memberikan kontribusi terbesar terhadap peningkatan akurasi (+0.028 AUC), diikuti pruning yang sangat efisien (-0.54 MB dengan degradasi AUC <0.8%), dan quantization yang nyaris tanpa kompromi performa (-2.38 MB, 0% loss AUC). Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan yang perlu diatasi pada tahap selanjutnya: (1) pendekatan berbasis frame statis belum sepenuhnya menangkap inkonsistensi temporal antar-frame, (2) resolusi input terbatas pada 224×224 piksel dapat mengurangi sensitivitas terhadap manipulasi sangat halus, (3) dataset yang digunakan didominasi wajah Barat sehingga representasi terhadap karakteristik wajah Indonesia masih terbatas, serta (4) pengukuran konsumsi daya hanya bersifat simulasi. Untuk mengatasi hal tersebut, pengembangan masa depan dapat mengintegrasikan modul temporal ringan (misalnya LSTM atau 3D-CNN efisien), menerapkan mekanisme multi-scale atau super-resolution input, memperkaya dataset dengan koleksi lokal seperti INA-FaceFake, serta melakukan pengujian langsung pada perangkat Android kelas menengah (Snapdragon 6/7 series) untuk mendapatkan metrik daya dan stabilitas real-world. Secara praktis, model ini siap diintegrasikan ke dalam aplikasi mobile mandiri atau plugin media sosial (TikTok, Instagram, WhatsApp) dengan antarmuka sederhana yang menampilkan peringatan instan seperti “Video ini terdeteksi sebagai deepfake (kepercayaan 99.3%)”. Implementasi ini akan memperkuat literasi digital masyarakat Indonesia, khususnya dalam menghadapi potensi penyalahgunaan deepfake pada pemilu dan isu publik sensitif. Penelitian ini juga memberikan kontribusi nyata terhadap Sustainable Development Goals, terutama SDG 9 (Industry, Innovation and Infrastructure) melalui pengembangan edge AI yang inklusif dan terjangkau, serta SDG 16 (Peace, Justice and Strong Institutions) dengan menyediakan alat efektif untuk memerangi disinformasi dan melindungi integritas informasi di ruang digital. Secara keseluruhan, model MobileViT-CBAM yang diusulkan merupakan solusi deteksi deepfake pertama di kelasnya yang berhasil menggabungkan akurasi tinggi (AUC 0.993), ukuran ultra-kompak (0.80 MB), dan kemampuan real-time (15.8 FPS) pada perangkat smartphone kelas menengah dan bawah, sehingga membuka jalan bagi verifikasi konten multimedia yang demokratis, mandiri, dan dapat diakses oleh masyarakat luas di Indonesia maupun negara berkembang lainnya.

Referensi

1. Cao, X., Zhong, P., Huang, Y., Huang, M., Huang, Z., Zou, T., & Xing, H. (2025). Research on Lightweight Algorithm Model for Precise Recognition and Detection of Outdoor Strawberries Based on Improved YOLOv5n. *Agriculture*, *15*(1), 90. https://doi.org/10.3390/agriculture15010090
2. Duanmu, A., Xue, S., Li, Z., Zhang, Y., & Ni, C. (2025). Rep-MobileViT: Texture and Color Classification of Solid Wood Floors Based on a Re-Parameterized CNN-Transformer Hybrid Model. *IEEE Access*, *13*, 39950–39963. https://doi.org/10.1109/ACCESS.2025.3545645

DOI: https://doi.org/10.31004/riggs.v4i4.3634

3. Jafar M Sidik. (2024, January 3). *Ancaman "Deepfake", pemanfaatan AI, dan Pemilu 2024*. Aceh.Antaranews.Com.
4. Lee, J., Kwon, Y., Park, S., Yu, M., Park, J., & Song, H. (2024). Q-HyViT: Post-Training Quantization of Hybrid Vision Transformers With Bridge Block Reconstruction for IoT Systems. *IEEE Internet of Things Journal*, *11*(22), 36384–36396. https://doi.org/10.1109/JIOT.2024.3403844
5. Liu, Y., Li, Y., Yi, X., Hu, Z., Zhang, H., & Liu, Y. (2022). Lightweight ViT Model for Micro-Expression Recognition Enhanced by Transfer Learning. *Frontiers in Neurorobotics*, *16*. https://doi.org/10.3389/fnbot.2022.922761
6. Liu, Y., Xing, H., & Hou, T. (2025). Sea Surface Floating Small-Target Detection Based on Dual-Feature Images and Improved MobileViT. *Journal of Marine Science and Engineering*, *13*(3), 572. https://doi.org/10.3390/jmse13030572
7. Palanisamy, B., Hassija, V., Chatterjee, A., Mandal, A., Chakraborty, D., Pandey, A., Chalapathi, G. S. S., & Kumar, D. (2025). Transformers for Vision: A Survey on Innovative Methods for Computer Vision. *IEEE Access*, *13*, 95496–95523. https://doi.org/10.1109/ACCESS.2025.3571735
8. Tahyudin, I., Prabuwono, A. S., Dianingrum, M., Pandega, D. M., Winarto, E., Nazwan, Rozak, R. 'Abdul, Lestari, P., & Tikaningsih, A. (2024). ResNet-CBAM in Medical Imaging: A High-Accuracy Tool for Stroke Detection from CT Scans. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 551–556. https://doi.org/10.1109/ICITISEE63424.2024.10730079
9. Teh, S., Sivakumar, S., & Motalebi, F. (2024). Vision Transformers for Biomedical Applications *. *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, 195–201. https://doi.org/10.1109/GECOST60902.2024.10474871
10. Tomás Lujambio. (2023, July 20). *Deepfakes Increase by 900% Annually: The Week in Cybersecurity*. Mexicobusiness.News.
11. Zhang, G., Li, W., Tang, Y., Chen, S., & Wang, L. (2025). Lightweight CNN-ViT with cross-module representational constraint for express parcel detection. *The Visual Computer*, *41*(5), 3283–3295. https://doi.org/10.1007/s00371-024-03602-0

DOI: https://doi.org/10.31004/riggs.v4i4.3634