CESS (Journal of Computing Engineering, System and Science) *e-ISSN: 2502-714x*
Vol. 9, No. 2, Juli 2024, pp.397-411 *p-ISSN: 2502-7131*

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: https://jurnal.unimed.ac.id/2012/index.php/cess

Komparasi Performa VGG19, ResNet50, DenseNet121 dan MobileNetV2 Dalam Mendeteksi Gambar Deepfake

Performance Comparation of VGG19, ResNet50, DenseNet121 and MobileNetV2 in Detecting Deepfake Image

Angeline[1], Harni Kusniyati[2*]

[1,2] Universitas Mercu Buana
Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta 11650
email: [1]41519120019@student.mercubuana.ac.id, [2]harni.kusniyati@mercubuana.ac.id

A B S T R A K

Deepfake secara pesat menjadi potensi ancaman keamanan siber yang dapat memanipulasi gambar, video, bahkan audio dengan sangat realistis sehingga manusia mengalami kesulitan dalam membedakan apakah sebuah media adalah asli atau merupakan hasil manipulasi kecerdasan buatan. CNN menjadi salah satu metode yang dikembangkan sebagai solusi. Banyaknya varian model CNN membuka potensi untuk pengembangan lebih lanjut. Penulis mengumpulkan dari berbagai sumber 1,000 citra wajah asli dan 1,000 citra wajah *deepfake* yang kemudian diperluas dengan teknik augmentasi data untuk melatih, memvalidasi, dan menguji empat varian model CNN yaitu VGG19, ResNet50, DenseNet121, dan MobileNetV2, dengan tujuan untuk menentukan varian yang paling efektif sebagai basis model yang dapat dikembangkan menjadi detektor *deepfake*. Evaluasi dan perbandingan performa dengan teknik *confusion matrix* menunjukkan bahwa di antara keempat model, ResNet50 memiliki performa terbaik dengan akurasi 91,5%, presisi 90%, dan recall 91,3%.

Kata Kunci: *Convolutional Neural Network; deepfake; klasifikasi gambar; VGG19; ResNet50; DenseNet121; MobileNetV2.*

A B S T R A C T

Deepfakes are rapidly becoming a potential cyber security threat that can manipulate images, videos and even audio so realistically that humans have difficulty distinguishing whether a piece of media is genuine or the result of manipulation by artificial intelligence. CNN is one of the methods developed as a solution. The large number of CNN variants opens up the potential for further development. The author collected from various sources 1,000 real facial images and 1,000 deepfake images which were then expanded using data augmentation technique to train, validate, and test four variants of the CNN model, namely VGG19,

**Penulis Korespondensi:*
email: harni.kusniyati@mercubuana.ac.id

ResNet50, DenseNet121, and MobileNetV2, with the aim of determining the most effective variant as a base model that can be developed as *deepfake* detector. Performance evaluation and comparison with confusion matrix technique shows that between the four models, ResNet50 performs best with 91,5% accuracy, 90% precision, and 91,3% recall.

Keywords: *Convolutional Neural Network; deepfake; image classification; VGG19; ResNet50; DenseNet121; MobileNetV2.*

1. PENDAHULUAN

Di tengah argumentasi mengenai kecanggihan dan kerugian yang dapat dibawakan oleh kecerdasan buatan (*artificial intelligence*), kemunculan *deepfake* menambah daftar di sisi negatif dari kemajuan pesat teknologi AI. *Deepfake* secara luas didefinisikan sebagai media yang dihasilkan menggunakan teknik AI dan *machine learning* [1], dimana AI dilatih untuk meniru media yang telah ada untuk 'menciptakan' karya baru. *Deepfake* hadir dalam berbagai rupa modalitas seperti gambar, video, dan audio [2] , dan terlepas dari potensi teknologi ini di bidang medis, edukasi, dan pariwisata [1], *deepfake* lebih dikenal karena potensi ancamannya terhadap berbagai bidang dan institusi baik pemerintahan, perusahaan, maupun masyarakat yang menggunakan serta bergantung pada sistem informasi digital.

Melalui kecerdasan buatan seperti *generative adversarial networks* (GAN), pengguna dapat mensintesis suara individu tertentu berdasarkan transkrip, menukar wajah seseorang ke tubuh orang lain dalam sebuah video, atau mensintesis video baru yang menampilkan seseorang mengucapkan hal tertentu berdasarkan audio yang disinkronkan ke wajah mereka [3].

Dalam dunia perbankan, deepfake dapat dengan mudah digunakan untuk menghindari penerapan prinsip KYC, memalsukan profil orang lain, menyalahgunakan privasi konsumen, dan bahkan membobol akun konsumen [4]. Dalam rana politik, deepfake dapat mendistorsi informasi yang digunakan pemilih dalam mengambil keputusan dan berpotensi merusak integritas pemilu yang demokratis [3]. Pada sektor media, AI memiliki potensi untuk menciptakan konten baru yang mungkin melanggar hak cipta seseorang [5] karena deepfake dapat berupa musik, artikel, atau bahkan film yang mirip dengan karya yang sudah ada.

Seiring berkembangnya AI, deepfake juga turut berkembang dan semakin sulit untuk diidentifikasi oleh mata manusia. Pada studi oleh Groh dkk. [6], 304 partisipan hanya dapat secara akurat mengidentifikasi deepfake dalam 66% percobaan. Eksperimen serupa oleh Kas dkk. [7], membuktikan bahwa meskipun partisipan memiliki kepercayaan diri tinggi akan kemampuan mereka dalam membedakan antara gambar wajah asli dan deepfake, 59 partisipan hanya dapat mengidentifikasi deepfake dengan benar pada 61% kasus. Di sisi lain, teknologi jaringan saraf terbukti memiliki akurasi tinggi dalam mengidentifikasi deepfake, seperti pada penelitian oleh Salman dan Naser [8] dengan model NN usulannya yang terstruktur dari 8 lapisan konvolusi dan pooling berhasil mendeteksi gambar wajah sintetik dengan akurasi pengujian 95%; Penelitian oleh St dkk. [9] yang mengimplementasikan histogram pola biner linier fisherface menggunakan teknik DBN classifier (FF-LBPH DBN) sebagai teknik pendeteksian gambar deepfake, dengan tingkat akurasi yang mencapai 98,82% terhadap dataset CASIA-WebFace, dan 97.82% untuk dataset DFFD; Model CNN yang dikembangkan Mu dkk. [10] berhasil dalam melakukan identifikasi antara wajah asli dengan wajah buatan deepfake dengan tingkat akurasi sebesar 91%. Dengan kemampuan superior

dalam mempelajari dan mengenali pola dari dataset besar, CNN menawarkan pendekatan berbasis kecerdasan buatan yang lebih efektif daripada metode tradisional [11].

Dengan mengangkat topik ini, penulis bertujuan untuk menguji empat varian model convolutional neural network (CNN) yaitu VGG, ResNet, DenseNet, dan MobileNet dalam membedakan antara gambar wajah asli dan wajah sintetik, dalam prosesnya mengevaluasi performa model untuk menemukan varian dengan performa terbaik agar dapat dikembangkan lebih lanjut sebagai alat deteksi deepfake.

2. DASAR/TINJAUAN TEORI

2.1. Penelitian Terdahulu

CNN merupakan salah satu algoritma deep learning yang lazim digunakan dalam pengklasifikasian citra, baik klasifikasi biner seperti pada penelitian deteksi penggunaan helm keselamatan [12] maupun klasifikasi multi-label seperti pada penelitian klasifikasi motif batik [13].

Sejauh ini terdapat beberapa penelitian yang mengaplikasikan deep learning untuk mendeteksi gambar *deepfake* yang dihasilkan oleh GAN. Salman dan Naser [8] dengan model usulannya yang terstruktur dari 8 lapisan konvolusi dan pooling berhasil mendeteksi gambar wajah sintetik dan asli dengan akurasi pengujian 95% dalam 150 epoch. Pada penelitian oleh St dkk. [9] , histogram pola biner linier fisherface menggunakan teknik DBN classifier (FF-LBPH DBN) diimplementasikan sebagai teknik pendeteksian gambar *deepfake*, dengan tingkat akurasi yang mencapai 98.82% terhadap dataset CASIA-WebFace, dan 97.82% untuk dataset DFFD. Shad dkk. [14] mengusulkan model *Visual Geometry Group* (VGG) yang mengimplementasikan arsitektur CNN untuk dilatih dan diuji dengan dataset Celeb-DF. Chang dkk. [15] merancang model CNN kustom dengan enam lapisan konvolusi yang masing-masing diikuti oleh lapisan BN dan *max-pooling*, fungsi aktivasi ReLu, dan *dropout*, serta fungsi padding untuk peningkatan presisi bagi kernel dalam pengecekan citra. Model rancangannya tidak bekerja dengan baik dan mencapai performa yang lebih buruk dibanding DenseNet, VGG, dan ResNet, tapi hasil perbandingan Shad menunjukkan bahwa VGGFace dapat mengenali gambar deepfake resolusi tinggi. Pendekatan inovatif lainnya oleh Raza dkk. [16] yang dinamai DFP (*deepfake predictor*) menggabungkan antara VGG16 dengan CNN dan dua lapisan Dense yang ditempatkan setelah lapisan Flatten, dan dilatih dengan dataset berisi 1,081 gambar wajah asli dan 960 gambar wajah palsu, dan mengungguli model-model lain seperti NAS-Net, Xception, MobileNet, dan VGG16. Dengan menembangkan DenseNet yang disematkan dengan fungsi aktivasi Leaky ReLU di setiap lapisan Dense, Patel dkk. [17] mendapatkan hasil yang menjanjikan pada tahap pelatihan dan validasi, namun mencapai akurasi 77% ketika diuji dengan dataset CelebDF yang memang dianggap kumpulan data yang menantang dalam deteksi deepfake. Zhang dkk. [18] mengusulkan TD-3DCNN yang merupakan modul 3D Inception yang diperkokoh Temporal Dropout untuk mendeteksi inkonsistensi pada tiap frame video, meningkatkan kemampuan representasi dan generalisasi model dan dibuktikan dengan AUC rata-rata 87.40%. Kohli dan Gupta [19] menguji sebuah model *frequency* CNN berlapis tiga yang dilatih dengan dataset FaceForensic++ dan Celeb-DF untuk mendeteksi pemalsuan wajah, menghasilkan sebuah model yang dapat bersaing dengan XceptionNet dan MesoNet. Sementara Abidin dkk. [20] memanfaatkan ResNext dan LSTM dalam deteksi video *deepfake* dan mencapai performa akurasi 90%, presisi 100%, dan recall 97% pada 60 frame. Guarnera dkk. [21] menggunakan algoritma *Expectation Maximization* (EM), memfokuskan model untuk mencari jejak konvolusi (*convolutional trace*) pada citra *deepfake*, berbeda dengan

pendekatan CNN yang mencari pola tersembunyi dari citra-citra yang dihasilkan GAN. PPCN yang diusulkan Liu dkk. [22] melakukan pendekatan berbeda melalui pemotongan citra menjadi bagian-bagian spesifik yang disebut *patches* seperti bagian mata, mulut, hidung, dan kulit, dan menyusun multi-tasking framework dimana cabang pertama mempelajari perbedaan antara patch dari citra wajah asli dan palsu, sementara cabang kedua menangkap inkonsistensi antara daerah wajah dan bukan wajah. Ketika PPCNN diuji, model menggabungkan hasil dari dua cabang untuk membuat keputusan global untuk menentukan apakah gambar tersebut asli atau palsu.

Penelitian-penelitian tersebut menunjukkan sebuah pola yaitu implementasi CNN dalam model mereka, baik sebagai arsitektur utama atau pendukung. CNN terbukti efektif dalam mengenali, mendeteksi, maupun mengklasifikasikan citra dengan mayoritas tingkat akurasi di atas 80%, bahkan dengan struktur CNN standar tanpa modifikasi.

Di antara penelitian-penelitian terdahulu yang telah dikaji oleh penulis, banyak model yang performanya bergantung pada dataset latih, dan pada umumnya sebuah dataset mengumpulkan gambar deepfake dari hanya satu tipe GAN. Karena itu, penulis mengadopsi ide Patel dkk. [17] yang menggabungkan beberapa dataset *deepfake* publik ke dalam suatu dataset baru, di dalamnya menampung citra *deepfake* dengan berbagai resolusi, kualitas, dan tingkat kerumitan identifikasi yang beragam. Hal ini diharapkan dapat meminimalisir bias model terhadap pola algoritma GAN tertentu.

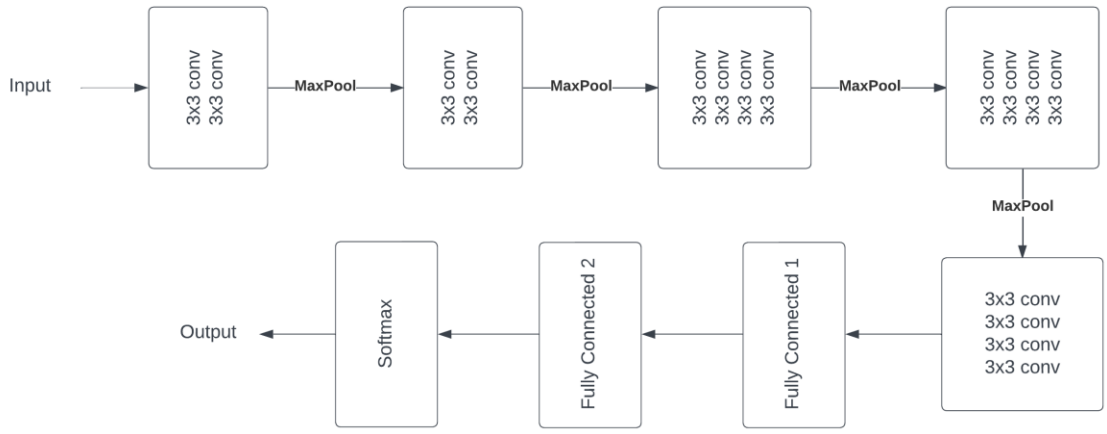
Para praktisi masih mencari algoritma ML maupun DL yang paling sesuai digunakan sebagai basis model detektor deepfake, terutama di antara varian kembangan CNN karena penerapannya yang dikhususkan untuk mempelajari data visual atau citra. Eksperimen yang akan dilaksanakan ini akan menentukan varian CNN mana yang paling efektif dari segi performa demi mendukung penelitian-penelitian selanjutnya di masa depan.

2.2. Teori Pendukung

2.2.1. Visual Geometry Group-Net (VGG-Net)

VGG dikembangkan oleh *Visual Geometry Group* di Universitas Oxford dalam penelitian oleh Karen Simonyan dan Andrew Zisserman yang berjudul "*Very Deep Convolutional Networks for Large-Scale Image Recognition*" pada tahun 2014. VGG terkenal karena kesederhanaan dan efektivitasnya. Dua model dari VGG, yaitu VGG16 yang terdiri dari 16 lapisan bobot dan VGG19 yang terdiri dari 19 lapisan bobot, tmenggunakan lapisan konvolusional 3x3 dengan ukuran filter kecil yang ditumpuk satu sama lain. VGG memanfaatkan kumpulan data ImageNet dan menggunakan jaringan saraf konvolusional dalam untuk mempelajari fitur hierarki dari gambar, sehingga memungkinkan klasifikasi objek dan pemandangan secara akurat.

VGG-Net mengikuti struktur dasar CNN namun dengan tambahan kedalaman lapisan. ‘Kedalaman’ mengacu pada jumlah lapisan konvolusi; VGG16 contohnya terbentuk atas 16 lapisan konvolusi, sementara VGG19 memiliki 19 lapisan konvolusi. VGG mengeksplorasi jaringan dengan kedalaman yang meningkat menggunakan arsitektur dengan filter konvolusi yang sangat kecil (3×3), yang menunjukkan bahwa peningkatan signifikan pada konfigurasi penemuan sebelumnya dapat dicapai dengan mendorong kedalaman ke bobot 16-19 lapisan (Zisserman & Simonyan, 2015). VGG menggunakan filter reseptif ukuran paling kecil yaitu 3x3. Blok penyusun dasar konfigurasi ini adalah tumpukan beberapa lapisan konvolusional dengan ukuran filter 3x3, stride 1, dan padding 1, diikuti oleh lapisan MaxPooling berukuran 2x2.

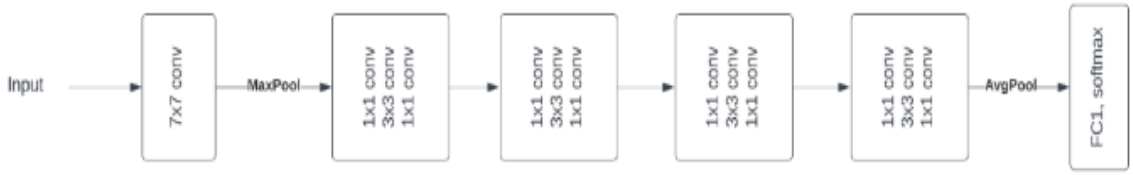


Gambar 1. Struktur lapisan VGG

2.2.2. Residual Neural Network (ResNet)

ResNet diperkenalkan oleh Kaiming He dkk. dari Microsoft Research dalam penelitian mereka "Deep Residual Learning for Image Recognition" pada tahun 2015. ResNet mengatasi masalah hilangnya gradien (vanishing gradient) di jaringan saraf dalam dengan memperkenalkan koneksi residual sehingga jaringan dapat mempelajari fungsi residual daripada secara langsung mempelajari fungsi pemetaan yang mendasarinya. Arsitektur ini memungkinkan pelatihan jaringan yang sangat dalam (hingga 152 lapisan) tanpa mengalami penurunan performa. ResNet telah dilatih sebelumnya pada kumpulan data ImageNet dan telah mencapai performa canggih dalam berbagai tugas pengenalan gambar karena kedalaman dan koneksi residual.

Residual Block yang menjadi keunikan ResNet tersusun dari koneksi-koneksi pintasan (identity shortcut connection) dimana setiap lapisan diumpankan ke lapisan jaringan berikutnya juga langsung ke lapisan berikutnya dengan melompati beberapa lapisan di antaranya. ResNet dianggap telah mencapai performa state-of-the-art (SOTA) bagi DNN, namun beberapa penelitian mendapati bahwa ResNet masih memiliki kelemahan seperti pembatasan kekuatan representasi jaringan yang diakibatkan oleh koneksi pintasan (Zhang et al., 2021).



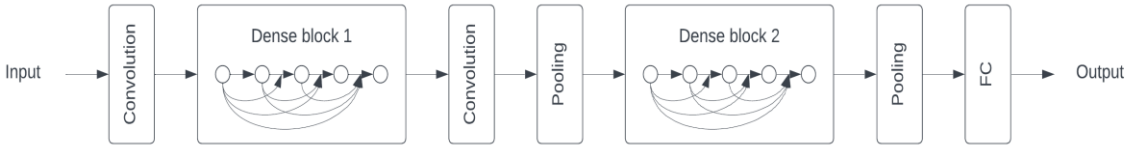
Gambar 2. Struktur lapisan ResNet

2.2.3. Dense Convolutional Network (DenseNet)

DenseNet yang diusulkan oleh para peneliti di Facebook AI Research (FAIR),memperkenalkan pola konektivitas baru antar lapisan. Salah satu varian populer, DenseNet121, merevolusi arus informasi dengan membangun koneksi langsung antara semua lapisan dalam satu blok padat. Tidak seperti CNN tradisional yang peta fiturnya digabungkan,

DenseNet menyebarkan peta fitur melalui penggabungan, memfasilitasi penggunaan kembali fitur, dan mendorong penyebaran fitur ke seluruh jaringan. DenseNet terdiri dari blok padat, lapisan transisi, dan pengumpulan rata-rata global, yang berpuncak pada lapisan yang terhubung sepenuhnya untuk klasifikasi. Pola konektivitas yang padat meningkatkan penggunaan kembali fitur, mengurangi redundansi parameter, dan mendorong penyebaran fitur, sehingga menghasilkan efisiensi parameter yang lebih baik dan peningkatan performa.

Pada DenseNet, setiap lapisan memperoleh masukan tambahan dari semua lapisan sebelumnya dan meneruskan peta fiturnya sendiri ke lapisan berikutnya, menyebabkan kondisi dimana setiap lapisan menerima “pengetahuan kolektif” dari semua lapisan sebelumnya. Karena hal ini, DenseNet memerlukan memori GPU yang besar dan seringkali waktu pelatihan yang lebih lama (C. Zhang et al., 2021). Pada DenseNet standar, komposisi struktur lapisan Dense Block terdiri atas Batch Normalization (BN), ReLU, dan lapisan konvolusi.



Gambar 3. Struktur lapisan DenseNet

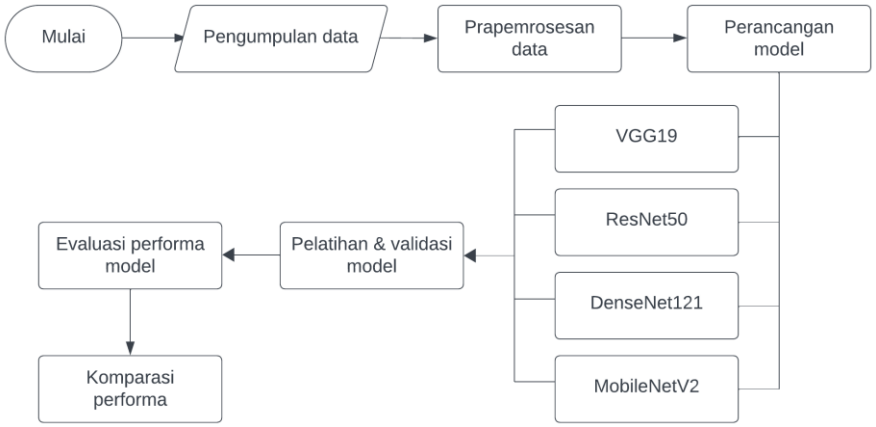
2.2.4. MobileNet

MobileNet yang dikembangkan oleh Google dirancang khusus untuk aplikasi visi seluler dan tertanam dengan sumber daya komputasi terbatas. Model ini diperkenalkan dalam jurnal “MobileNetV2: Inverted Residuals and Linear Bottleneck" oleh Mark Sandler dkk. pada tahun 2018. MobileNet menggunakan konsep residual block terbalik dengan linear bottleneck untuk mencapai keseimbangan yang baik antara akurasi dan efisiensi. Arsitekturnya menggabungkan konvolusi yang dapat dipisahkan secara mendalam untuk mengurangi jumlah parameter dan biaya komputasi sekaligus menjaga kapasitas representasi. MobileNet telah dilatih sebelumnya pada kumpulan data ImageNet dan telah diadopsi secara luas di perangkat seluler untuk tugas-tugas seperti klasifikasi gambar, deteksi objek, dan segmentasi semantik.

Diciptakan dengan tujuan untuk dapat digunakan pada perangkat seluler, konsep dasar di balik MobileNet adalah untuk membagi proses konvolusi menjadi depthwise convolution—yang menerapkan filter tunggal ke setiap saluran masukan—dan pointwise convolution—sebuah filter konvolusi berukuran 1x1 yang menggabungkan keluaran dari proses konvolusi depthwise. Ketika konvolusi standar memfilter dan menggabungkan masukan menjadi serangkaian keluaran baru dalam satu langkah, depthwise convolution membaginya menjadi dua lapisan, yaitu lapisan terpisah untuk pemfilteran dan lapisan terpisah untuk penggabungan. Faktorisasi ini berdampak pada pengurangan komputasi dan ukuran model secara drastis.

3. METODE

3.1. Tahap Penelitian



Gambar 4. Alur kerja tahapan penelitian

Penelitian dilaksanakan tahap per tahap dari pengumpulan data, pra pemrosesan data, perancangan empat model (VGG19, ResNet50, DenseNet121, MobileNetV2) yang akan melalui pelatihan dan validasi, evaluasi performa, hingga komparasi hasil performa keempat model. Pada proses pengumpulan data, penulis mengumpulkan data berupa citra wajah manusia dan *deepfake* dari beberapa sumber di Kaggle. Dataset yang dikumpulkan terdiri dari gambar-gambar dalam berbagai ukuran. Untuk dapat melakukan pemrosesan data dengan akurat, semua gambar diubah ukurannya melalui metode *down sampling* menjadi 256x256 untuk mengurangi resolusi spasial dan menghemat sumber daya. Data citra yang telah diseragamkan ukurannya kemudian dikonversi ke tipe array dan diberi label yang merepresentasikan kelas data dengan nilai 0 untuk *Deepfake* dan 1 untuk *Human*. Setelah dilabel, data dibagi ke dalam tiga subset yaitu train, test, dan validation dengan proporsi rasio 80:10:10. Terdapat empat model yang akan dilatih yaitu VGG19, ResNet50, DenseNet121, dan MobileNetV2. Keempat model tersebut tersedia dalam library TensorFlow Keras dan dapat diimpor ke lingkungan pemrograman, dalam kasus ini Google Collab. Setelah melalui proses perancangan, model akan dilatih untuk mengenali citra wajah asli maupun *deepfake* yang telah disiapkan dalam *training dataset* (dataset latih). Tahap validasi diperlukan untuk mengevaluasi performa model selama masa pelatihan guna penyetelan atau penyesuaian hyperparameter lebih lanjut dan menggunakan dataset validasi. Untuk pengujian, digunakan *testing dataset* (dataset uji) sebagai penilaian akhir performa model yang sudah dirancang. Tahap evaluasi menggunakan teknik Confusion Matrix dilakukan untuk mengevaluasi performa model-model yang telah dilatih dan diuji. Performa antar model yang dikembangkan dalam penelitian ini kemudian akan dikomparasikan untuk menentukan model dengan kinerja terbaik.

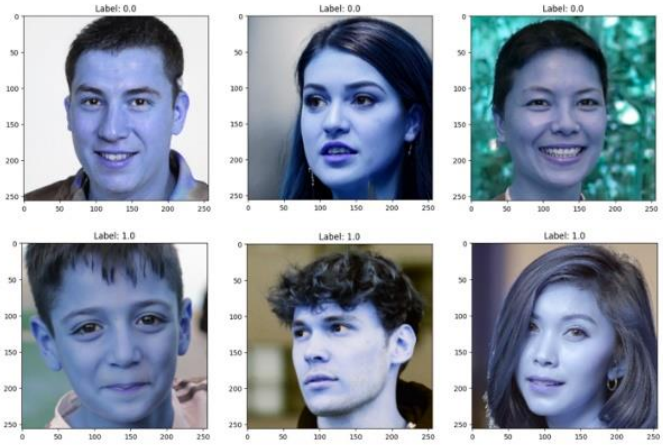
3.2. Dataset

Dari keseluruhan 2.000 data dalam dataset yang digunakan, 1.000 merupakan citra wajah asli dan 1.000 merupakan citra wajah sintetik. Data dibagi menjadi dataset latih, dataset uji, dan dataset validasi dengan rasio 80:10:10 menggunakan fungsi *train_test_split()* yang membagi 1.600 data dalam dataset latih, 200 data dalam dataset uji, dan 200 data dalam dataset validasi yang dipilih dan dipisahkan secara acak.

3.3. Pra pemrosesan

Data citra yang dikumpulkan dalam dataset memiliki dimensi panjang dan lebar yang bervariasi, maka dari itu dilakukan resizing pada tahap awal pra pemrosesan. Setiap citra diubah ukurannya menjadi 256x256 menggunakan fungsi *cv2.resize().* Data berformat citra dikonversi ke array, pada prosesnya ditambahkan satu dimensi bernama 'label' untuk identifikasi kelas data tersebut; label 0 mewakili *deepfake* sementara label 1 mewakili citra wajah asli. Augmentasi data dilakukan untuk memperluas ukuran dataset secara artifisial dengan membuat data-data baru melalui transformasi geometri dan spasi warna (flipping, resizing, cropping, kecerahan, kontras) pada dataset yang ada untuk meningkatkan ukuran dan keragaman dari dataset latih.

Menggunakan fungsi *plt.imshow(),* data array citra divisualisasikan melalui grafik plot. Data berlabel 0 merupakan citra wajah sintetik sementara data berlabel 1 merupakan citra wajah asli.



Gambar 5. Sampel citra wajah manusia dan deepfake

3.4. Perancangan Model

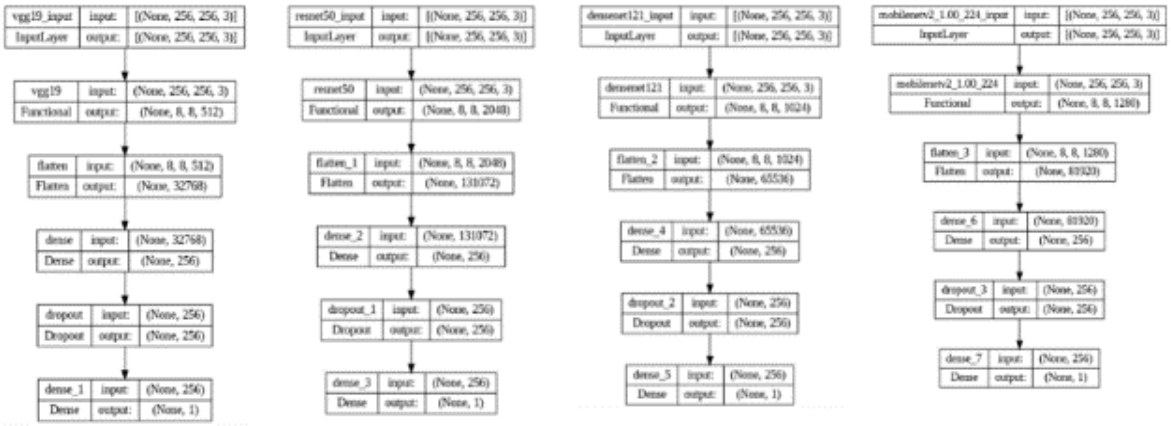
Keempat model merupakan model pretrained atau telah dilatih menggunakan database ImageNet dengan tujuan untuk memaksimalkan performa dari model yang telah memiliki serangkaian bobot dan bias awal untuk pemenuhan tugas yaitu mengidentifikasi gambar wajah deepfake. Tujuan penggunaan model pretrained adalah untuk menghemat waktu dan sumber daya. Hyperparameter yang digunakan adalah:

Tabel 1. Hyperparameter model

Hyperparameter	Nilai
optimizer	*adam*
epoch	11
batch_size	32
loss	*Binary_crossentropy*
metrics	*accuracy*

Untuk penelitian ini, digunakan arsitektur VGG19, ResNet50, DenseNet121, dan MobileNetV2 *pretrained* yang diimpor dari *library tensorflow* dan telah memiliki bobot dan bias pelatihan dari ImageNet. Di atas model *pretrained*, ditambahkan lapisan flatten, lapisan dense, lapisan dropout, dan lapisan dense output untuk tugas klasifikasi. Lapisan input dari

setiap model menerima input berupa citra berukuran 256 × 256 dan warna RGB yang pada prosesnya mengalami pengurangan dimensi spasial menjadi 8 × 8 dan kedalaman sesuai dengan jumlah *filter* di lapisan konvolusi akhir tiap model.

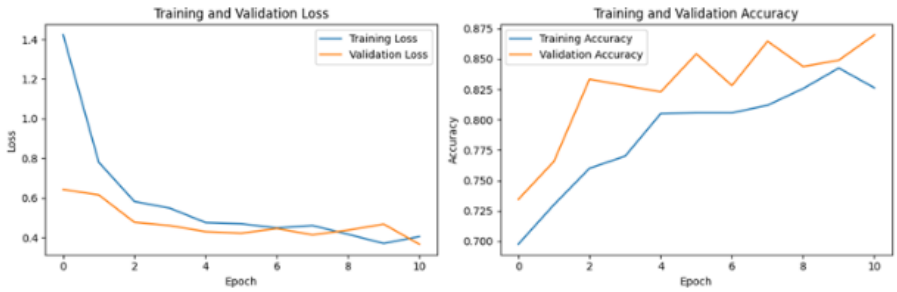


Gambar 6. Plot arsitektur model

4. HASIL DAN PEMBAHASAN

Setiap model melewati 11 iterasi dan 50 steps per iterasi dengan ukuran batch 32. Dilakukan penghitungan data loss dan data accuracy setiap model. Training loss adalah metrik yang digunakan untuk menilai kesesuaian model dengan data pelatihan, sementara validation loss mengukur kesesuaian model dengan data validasi. Dalam deep learning, loss adalah jarak perbedaan antara label yang sebenarnya dan label prediksi yang berusaha diminimalkan oleh jaringan saraf. Untuk meminimalkan jarak ini, jaringan saraf belajar dengan cara menyesuaikan bobot dan bias. Training accuracy merupakan nilai dari penghitungan akurasi dari prediksi model terhadap dataset latih, sementara validation accuracy adalah nilai penghitungan akurasi dari prediksi model terhadap dataset validasi. Keempat metrik ini memberikan gambaran performa dan kecocokan model terhadap dataset latih dan dataset validasi.

4.1. Loss dan Akurasi



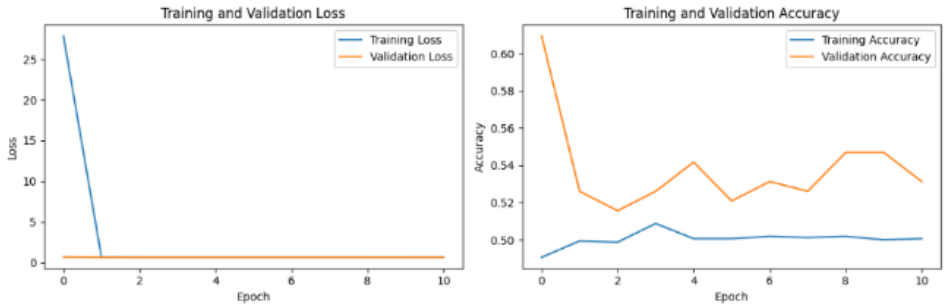
Gambar 7. Grafik training-validation loss dan accuracy VGG19

Tingkat loss diukur untuk dataset latih dan dataset validasi. Terhadap dataset training, nilai loss mengalami penurunan signifikan setelah iterasi ke-1 dan penurunan berkala di setiap iterasi selanjutnya hingga mencapai di bawah 0.4 pada iterasi ke-10. Nilai loss terhadap dataset validasi cenderung stabil setelah iterasi ke-2 dan bergerak mendekati 0 terkecuali pada iterasi ke-9 dimana nilai loss mengalami peningkatan namun sedikit meningkat di akhir hingga berada pada nilai yang lebih rendah dibandingkan training loss.

Akurasi model terhadap dataset latih terus mengalami peningkatan di setiap iterasinya dengan peningkatan yang melebihi 82.5%, namun mengalami penurunan pada iterasi ke-9. Sementara untuk data validasi, grafik akurasi yang didapatkan mengalami inkonsistensi dimana grafik meningkat dan kemudian menunjukkan beberapa fluktuasi, mencapai puncaknya di awal iterasi ke-3, dan kemudian mengalami penurunan dan peningkatan signifikan. Dari plot tersebut dapat disimpulkan bahwa model VGG19 berkinerja baik. Penurunan yang terus-menerus pada training loss dan validation loss menunjukkan bahwa model tidak mengalami overfitting, dan tren peningkatan training accuracy dan validation accuracy menunjukkan bahwa model tersebut dapat menggeneralisasi data validasi dengan baik

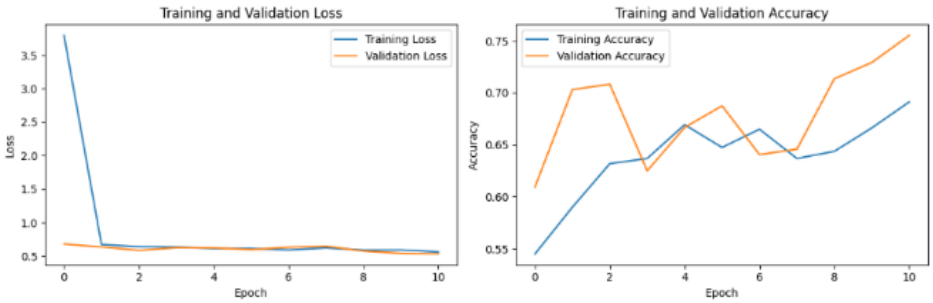
Gambar 8. Grafik training-validation loss dan accuracy ResNet50

Grafik training and validation loss pada gambar 4.3 menunjukkan tingkat loss yang sangat stabil terhadap dataset latih maupun dataset validasi di angka yang rendah, yang menunjukkan bahwa model telah mempelajari dataset secara efektif tanpa mengalami overfitting. Dari segi akurasi, model bereaksi sangat baik pada dataset validasi di awal, melebihi dataset latih, namun peningkatan training accuracy berkala menunjukkan pembelajaran model berlangsung baik.



Gambar 9. Grafik training-validation loss dan accuracy DenseNet121

Penurunan signifikan pada training loss hingga mendekati nol dari iterasi ke-1 mengindikasikan bahwa DenseNet121 mengalami overfitting. Validation loss berada di angka nol dari awal iterasi dan sepenuhnya tidak menunjukkan perubahan, sementara validation accuracy justru menunjukkan tren penurunan, yang menunjukkan instabilitas performa model. Training accuracy cenderung konstan di angka 0,5 tanpa mengalami peningkatan bahkan hingga iterasi ke-10.



Gambar 10. Grafik training-validation loss dan accuracy MobileNetV2

Setelah penurunan awal pada iterasi ke-1, kedua kurva loss menjadi stabil, yang menunjukkan bahwa model tidak mengalami overfitting berlebihan, meskipun validation loss menunjukkan lebih banyak fluktuasi dibandingkan training loss yang mengindikasikan adanya variabilitas dalam performa MobileNetV2 terhadap dataset validasi. Validation accuracy mengalami fluktuasi hingga berada pada angka yang lebih tinggi dibandingkan training accuracy pada akhir pelatihan, yang menunjukkan bahwa model dapat menggeneralisasi data dengan baik

4.2. Confusion Matrix

Hasil berupa ukuran performa tiap model disajikan secara kuantitatif menggunakan teknik Confusion Matrix. *Confusion matrix* merupakan teknik acuan untuk mengevaluasi performa machine learning. *True Positive* (TP) mewakili data positif yang diprediksi dengan benar, *True Negative* (TN) mewakili data negatif yang diprediksi dengan benar, *False Positive* (FP) mewakili data positif yang salah diprediksi sebagai negatif oleh model, dan *False Negative* (FN) merupakan data negatif yang salah diprediksi oleh model sebagai positif. Dari elemen-elemen pada Confusion Matrix tersebut dapat ditentukan nilai akurasi, presisi, recall, dan skor F1.

Tabel 2. Elemen confusion matrix

		Kondisi Prediksi	
	Total = P + N	Positif (P)	Negatif (N)
Kondisi Aktual	Positif (P)	True Positive (TP)	False Negative (FN)
	Negatif (N)	False Positive (FP)	True Negative (TN)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Akurasi merepresentasikan jumlah data yang diklasifikasikan dengan benar (True Positive) dibandingkan jumlah total data.

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

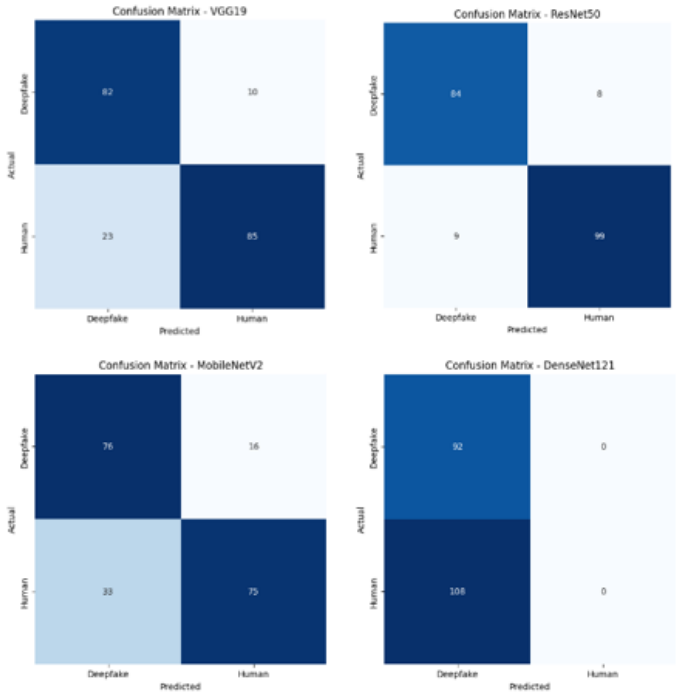
$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Skor\ F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (4)$$

Skor F1 mengacu pada perbandingan rata-rata presisi dan recall yang dibobotkan.

Confusion matrix pada penelitian ini merupakan fungsi yang diimpor dari library sklearn dan dikombinasikan dengan fungsi figure dari library matplotlib. Pada grafik confusion matrix yang ditampilkan pada gambar 4.2, 4.3, 4.4, dan 4.5, axis y mewakili label aktual, yaitu label sebenarnya pada data, sementara axis x mewakili nilai prediksi, yaitu label pada data yang diprediksi oleh model.



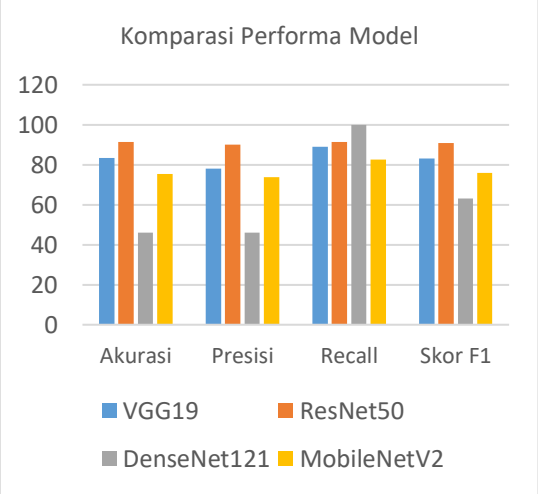
Gambar 11. Hasil confusion matrix

4.3. Komparasi Performa

Tabel 3. Komparasi performa model

Model	Akurasi	Presisi	Recall	Skor F1
VGG19	0.835	0.78	0.891	0.832
ResNet50	0.915	0.90	0.913	0.908
DenseNet121	0.46	0.46	1	0.63
MobileNetV2	0.755	0.737	0.826	0.76

Tabel 4. Grafik komparasi performa model



5. KESIMPULAN

Tiga dari empat model yang diuji menunjukkan tingkat akurasi yang lebih baik dalam mengenali dan mendeteksi gambar wajah deepfake dibandingkan peserta manusia dalam penelitian Kas dkk. (2020) dan Bray dkk. (2023) terkecuali DenseNet121 yang hanya mencapai akurasi 46%. Dari hasil ini dapat disimpulkan bahwa beberapa jaringan saraf dapat melampaui akurasi manusia dalam mendeteksi *deepfake*. ResNet50 menghasilkan performa terbaik dalam mendeteksi gambar *deepfake*, terbukti dari keunggulan nilai akurasi, presisi, dan skor F1 dibandingkan dengan VGG19, DenseNet121, dan MobileNetV2. Meskipun recall ResNet50 lebih rendah dibandingkan DenseNet121, metrik performa lain dari DenseNet121 menunjukkan kinerja buruk sehingga dapat disimpulkan bahwa secara keseluruhan ResNet50 tetap lebih unggul. Arsitektur CNN khususnya ResNet50 memiliki potensi untuk dipergunakan sebagai alat deteksi deepfake, namun pada prakteknya masih memiliki akurasi yang belum tergolong baik jika tanpa pengembangan atau tuning lanjutan. Selain ResNet50, perlu juga dievaluasi arsitektur deep learning lain yang berkemampuan untuk mengenali dan mengklasifikasi pola pada citra.

REFERENSI

[1] Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J., & Russel-Benner, R. (2023). Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation*, *125*.

[2] Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023c). Testing human ability to detect 'deepfake' images of human faces. *Journal of Cybersecurity*, *9*(1), tyad011. https://doi.org/10.1093/cybsec/tyad011

[3] Diakopoulos, N., & Johnson, D. (2021b). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

[4] Gunawan, I. J., & Janisriwati, S. (2023b). Legal Analysis on the Use of Deepfake Technology: Threats to Indonesian Banking Institutions. *Law and Justice*, *8*(2), 192–210. https://doi.org/10.23917/laj.v8i2.2513

[5] Akbari, R. N., & Fithry, A. (2024b). MENGANALISIS PENGARUH HAK CIPTA DALAM GANGGUAN AI PADA SEKTOR MEDIA. *Prosiding SNAPP : Sosial Humaniora, Pertanian, Kesehatan dan Teknologi*, *2*(1), 377–383. https://doi.org/10.24929/snapp.v2i1.3159

[6] Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022b). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, *119*(1), e2110013119. https://doi.org/10.1073/pnas.2110013119

[7] Kas, S., Hes, T., Jansen, B., & Post, R. (n.d.-b). *Do you know if I'm real? An experiment to benchmark human recognition of AI-generated faces*.

[8] Salman, F. M., & Abu-Naser, S. S. (2022b). *Classification of Real and Fake Human Faces Using Deep Learning*. *6*(3).

[9] St, S., Ayoobkhan, M. U. A., V, K. K., Bacanin, N., K, V., Štěpán, H., & Pavel, T. (2022b). Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*, *8*, e881. https://doi.org/10.7717/peerj-cs.881

[10] Mu, J., Adrezo, M., & Haikal, A. N. (2024b). Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network. *Teknika*, *13*(1), 45–50. https://doi.org/10.34148/teknika.v13i1.705

[11] Bachri, C. M., & Gunawan, W. (2024). *Deteksi Email Spam menggunakan Algoritma Convolutional Neural Network (CNN)*. *10*(1).

[12] Mianah, A. N., Diah Arie Widhining K, & Farrady Alif Fiolana. (2023b). Klasifikasi Helm Keselamatan Mengunakan Metode Convolutional Neural Network (CNN). *JOURNAL ZETROEM*, *5*(2), 94–102. https://doi.org/10.36526/ztr.v5i2.2765

[13] Bariyah, T., Rasyidi, M. A., & Ngatini, N. (2021b). Convolutional Neural Network untuk Metode Klasifikasi Multi-Label pada Motif Batik. *Techno.Com*, *20*(1), 155–165. https://doi.org/10.33633/tc.v20i1.4224

[14] Shad, H. S., Rizvee, Md. M., Roza, N. T., Hoq, S. M. A., Monirujjaman Khan, M., Singh, A., Zaguia, A., & Bourouis, S. (2021b). Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network. *Computational Intelligence and Neuroscience*, *2021*, 1–18. https://doi.org/10.1155/2021/3111676

[15] Chang, X., Wu, J., Yang, T., & Feng, G. (2020b). DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network. *2020 39th Chinese Control Conference (CCC)*, 7252–7256. https://doi.org/10.23919/CCC50068.2020.9189596

[16] Raza, A., Munir, K., & Almutairi, M. (2022b). A Novel Deep Learning Approach for Deepfake Image Detection. *Applied Sciences*, *12*(19), 9820. https://doi.org/10.3390/app12199820

[17] Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F. (2023b). An Improved Dense CNN Architecture for Deepfake Image Detection. *IEEE Access*, *11*, 22081–22095. https://doi.org/10.1109/ACCESS.2023.3251417

[18] Zhang, D., Li, C., Lin, F., Zeng, D., & Ge, S. (2021b). Detecting Deepfake Videos with Temporal Dropout 3DCNN. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 1288–1294. https://doi.org/10.24963/ijcai.2021/178

[19] Kohli, A., & Gupta, A. (2021b). Detecting DeepFake, FaceSwap and Face2Face facial forgeries using frequency CNN. *Multimedia Tools and Applications*, *80*(12), 18461–18478. https://doi.org/10.1007/s11042-020-10420-8

[20] Abidin, M. I., Nurtanio, I., & Achmad, A. (2022b). Deepfake Detection in Videos Using Long Short-Term Memory and CNN ResNext. *ILKOM Jurnal Ilmiah*, *14*(3), 178–185. https://doi.org/10.33096/ilkom.v14i3.1254.178-185

[21] Guarnera, L., Giudice, O., & Battiato, S. (2020b). Fighting Deepfake by Exposing the Convolutional Traces on Images. *IEEE Access*, *8*, 165085–165098. https://doi.org/10.1109/ACCESS.2020.3023037

[22] Liu, J., Zhu, K., Lu, W., Luo, X., & Zhao, X. (2021b). A lightweight 3D convolutional neural network for deepfake detection. *International Journal of Intelligent Systems*, *36*(9), 4990–5004. https://doi.org/10.1002/int.22499