

BAB II

KAJIAN LITERATUR

2.1 Hoaks Visual dan Media Sosial

Kata *Hoax* berasal dari bahasa Inggris yang artinya tipuan, menipu, berita bohong, berita palsu atau kabar burung yang disebar oleh seseorang. Jadi dapat dikatakan bahwa hoaks adalah kata yang berarti ketidakbenaran sebuah informasi. Hoaks bukan singkatan tetapi satu kata dalam bahasa Inggris yang punya arti sendiri. Sedangkan definisi hoaks menurut wikipedia adalah: “Sebuah pemberitaan palsu adalah usaha untuk menipu atau mengakali pembaca/pendengarnya agar mempercayai sesuatu, padahal sang pencipta berita palsu tersebut tahu bahwa berita tersebut palsu dan tidak benar”. Di era internet ini, masyarakat bebas menyampaikan pendapat atau opini nya, baik melalui lisan, media cetak, maupun media elektronik/*online* [2]. Hoaks visual itu sendiri diartikan sebagai suatu informasi dalam bentuk gambar atau video yang sengaja diubah, dipotong konteksnya, untuk menyesatkan penerima [3].

Hoaks dapat menimbulkan dampak yang sangat merugikan, baik bagi individu maupun masyarakat secara keseluruhan. Informasi palsu dapat menimbulkan kepanikan, ketidakpercayaan, dan bahkan dapat memicu tindakan yang merugikan. Oleh karena itu, penting bagi masyarakat untuk selalu memeriksa kebenaran informasi sebelum membagikannya atau mengambil tindakan berdasarkan informasi tersebut, dengan literasi digital masyarakat dapat memahami dan mengelola informasi yang tersebar di media sosial. Literasi di era digital ini sangat penting untuk dibahas, karena masyarakat harus memiliki kemampuan membedakan informasi yang benar dan hoaks dalam media sosial. Tetapi masalah yang dihadapi adalah kurangnya literasi di era digital di kalangan masyarakat. Banyak masyarakat yang tidak memahami bagaimana cara membedakan informasi yang benar dan hoaks di media sosial. Hal ini dapat menyebabkan masyarakat mudah terpengaruh oleh hoaks yang dapat membahayakan diri sendiri dan orang lain. Selain itu, penggunaan media sosial yang tidak bijak juga dapat menyebabkan masyarakat kehilangan waktu yang berharga dan mengalami gangguan kesehatan mental [4].

Lahirnya media sosial menjadikan pola perilaku masyarakat mengalami pergeseran baik budaya, etikan dan norma yang ada. Indonesia dengan jumlah penduduk yang besar dengan berbagai kultur suku, ras dan agama yang beraneka ragam memiliki banyak sekali potensi perubahan sosial. Dari berbagai kalangan dan usia hampir semua masyarakat Indonesia memiliki dan menggunakan media sosial sebagai salah satu sarana guna memperoleh dan menyampaikan informasi ke publik [14].

2.1.1 Hoaks Visual

Hoaks visual merupakan bentuk informasi palsu yang disajikan dalam format gambar atau video yang telah dimanipulasi untuk menyesatkan persepsi publik. Berbeda dengan hoaks berbasis teks, hoaks visual memiliki tingkat kepercayaan yang lebih tinggi karena manusia cenderung lebih mudah mempercayai informasi dalam bentuk visual. Perkembangan teknologi, khususnya kecerdasan buatan, semakin mempermudah pembuatan konten visual palsu yang tampak realistis dan sulit dikenali secara langsung. Hal ini menyebabkan hoaks visual menjadi semakin berbahaya karena dapat mempengaruhi opini publik, menimbulkan kesalahpahaman, serta berdampak pada berbagai aspek kehidupan termasuk sosial dan ekonomi [15]. Seperti di Gambar 2.1 yaitu sebuah berita sekaligus konten visual berupa gambar yang menampilkan Presiden Rusia Vladimir Putin terkena serangan jantung atau jatuh tersungkur tersebut merupakan informasi palsu (hoaks) [15].



Gambar 2. 1 Contoh Hoaks Visual di Media Sosial[15].

Dampak dari hoaks visual tidak hanya terbatas pada kesalahan informasi, tetapi juga dapat menimbulkan konsekuensi yang lebih luas seperti pembentukan opini publik yang keliru, konflik sosial, hingga kerugian pada sektor ekonomi dan bisnis. Informasi visual yang dimanipulasi dapat merusak reputasi individu maupun organisasi, serta menurunkan tingkat kepercayaan masyarakat terhadap informasi digital secara keseluruhan. Oleh karena itu, hoaks visual menjadi salah satu tantangan utama dalam era digital yang memerlukan perhatian khusus, baik dari sisi teknologi deteksi maupun peningkatan literasi digital masyarakat.

Selain itu, penyebaran hoaks visual sangat dipengaruhi oleh karakteristik media sosial yang mengutamakan kecepatan dan interaksi pengguna. Algoritma media sosial cenderung mempromosikan konten yang menarik perhatian, termasuk gambar dan video yang bersifat sensasional, sehingga hoaks visual dapat dengan cepat menjadi viral tanpa melalui proses verifikasi yang memadai. Kondisi ini diperparah oleh rendahnya literasi digital masyarakat, di mana banyak pengguna cenderung langsung membagikan informasi tanpa melakukan pengecekan kebenaran terlebih dahulu [16].

2.1.2 Media Sosial

Media sosial merupakan platform digital yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, serta memproduksi konten secara mandiri dalam jaringan yang luas dan terbuka. Perkembangan media sosial telah mengubah pola komunikasi masyarakat dari yang sebelumnya bersifat satu arah menjadi dua arah bahkan multi-arah, sehingga setiap individu dapat berperan sebagai produsen sekaligus konsumen informasi. Kemudahan akses serta kecepatan distribusi informasi menjadikan media sosial sebagai sarana utama dalam penyebaran informasi di era digital, baik informasi yang bersifat benar maupun yang mengandung disinformasi [17].

Karakteristik media sosial yang menekankan kecepatan dan interaksi pengguna juga menimbulkan tantangan serius, khususnya dalam penyebaran hoaks. Algoritma media sosial cenderung memprioritaskan konten yang menarik perhatian dan memiliki tingkat interaksi tinggi, sehingga informasi yang bersifat sensasional termasuk hoaks lebih mudah tersebar luas dibandingkan informasi yang telah terverifikasi. Selain itu, rendahnya literasi digital masyarakat menjadi faktor utama yang mempercepat penyebaran hoaks, di mana banyak pengguna cenderung membagikan informasi tanpa melakukan verifikasi terlebih dahulu [16].

Dampak dari fenomena ini tidak hanya terbatas pada penyebaran informasi yang salah, tetapi juga dapat mempengaruhi opini publik, menimbulkan kepanikan, serta mengganggu stabilitas sosial dan ekonomi. Dalam konteks bisnis, media sosial juga dapat menjadi sarana penyebaran informasi negatif yang berpotensi merusak reputasi perusahaan dan menurunkan kepercayaan konsumen. Oleh karena itu, pemahaman terhadap karakteristik dan dampak media sosial menjadi penting sebagai dasar dalam mengembangkan metode deteksi dan pengendalian konten manipulatif, termasuk *deepfake*.

2.2 Deepfake

Pernahkah kita membayangkan sebuah realitas digital di mana batasan antara rekaman asli dan simulasi buatan menjadi hampir tidak terlihat sama sekali? Di era transformasi informasi yang begitu masif ini, kita sering kali menjumpai konten visual maupun audio yang tampak sangat meyakinkan, namun sebenarnya merupakan hasil rekayasa algoritma yang sangat canggih [18]. Fenomena ini muncul sebagai hasil dari evolusi kecerdasan buatan yang mampu memetakan identitas seseorang ke dalam medium lain dengan tingkat presisi yang mencengangkan, memicu perdebatan luas mengenai integritas data di ruang siber.

Dalam Teknologi yang mendasari fenomena ini bekerja dengan cara mempelajari pola wajah, suara, dan gerak-gerik unik dari ribuan sampel data untuk menciptakan sebuah tiruan yang identik. Di balik kemampuannya yang memukau sekaligus meresahkan, terdapat mekanisme pembelajaran mesin mendalam yang secara konstan membandingkan hasil sintesis dengan data aslinya hingga mencapai tahap kemiripan yang sulit dibedakan oleh mata telanjang [18]. Secara terminologi, inovasi yang menggabungkan kecanggihan kecerdasan buatan dengan manipulasi konten ini dikenal *Deepfake*.

2.2.1 Definisi Deepfake

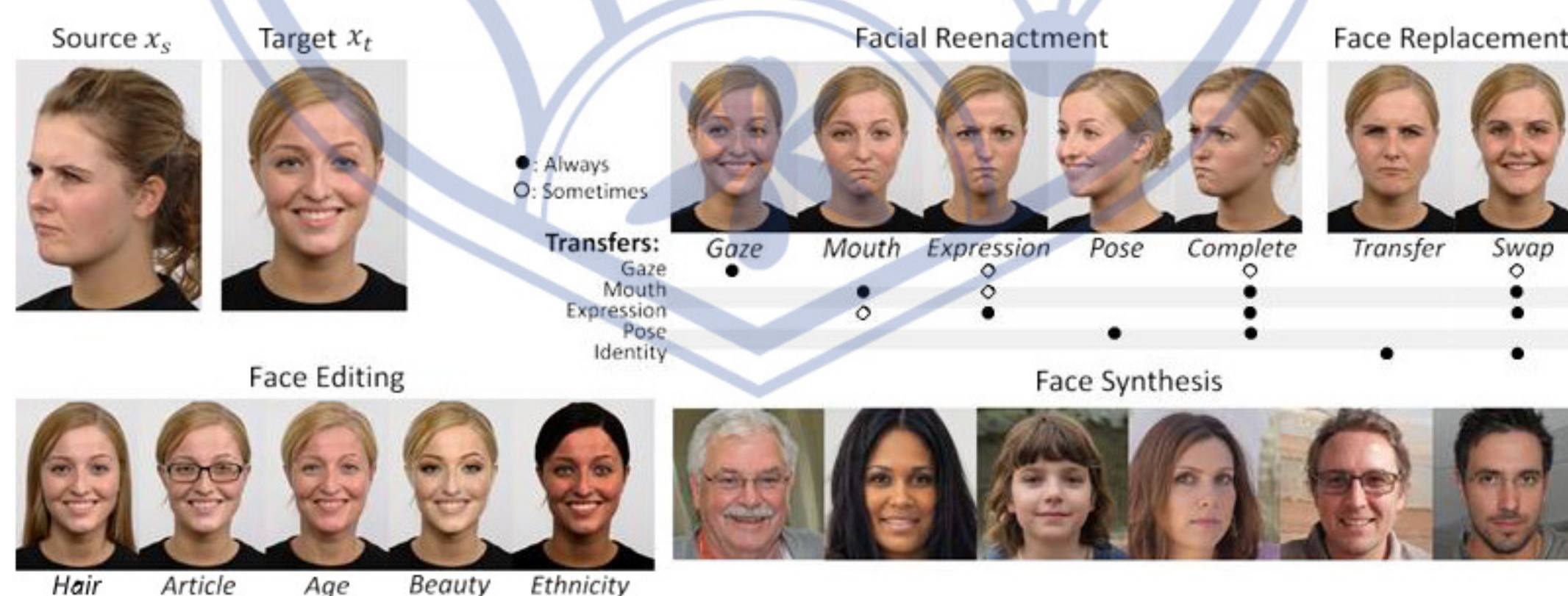
Deepfake secara harfiah merupakan gabungan dari istilah *deep learning* dan *fake*. Secara teknis, ini adalah media sintesis (gambar, audio, atau video) di mana citra seseorang digantikan dengan kemiripan orang lain menggunakan jaringan saraf tiruan yang dikenal sebagai *Generative Adversarial Networks* (GANs) [19]. Bayangkan sebuah sistem komputer yang terdiri dari dua "pemain": satu bertugas membuat tiruan wajah yang palsu (generator), dan satu lagi bertugas mendeteksi apakah wajah itu palsu (diskriminator). Mereka terus bertarung hingga sang generator mampu menciptakan hasil yang begitu sempurna sehingga sang diskriminator pun tidak bisa lagi membedakan mana yang asli dan mana yang buatan [19].

Visualisasi umum dari teknologi ini sering kita temukan dalam video di media sosial, seperti video viral Tom Cruise yang melakukan aksi sulap atau menari, padahal video tersebut sepenuhnya adalah hasil rekayasa oleh akun *@deeptomcruise*. Data menunjukkan bahwa penyebaran konten semacam ini meningkat sangat pesat; laporan dari sebuah perusahaan keamanan AI mencatat bahwa pada tahun 2019 saja, terdapat lebih dari 14.678 video *deepfake* yang beredar secara daring, dan angka ini terus melonjak ribuan persen seiring kemudahan akses aplikasi berbasis AI [20]. Fenomena ini bukan lagi sekadar

eksperimen laboratorium, melainkan alat yang mampu mengubah realitas visual dengan hanya bermodalkan foto wajah target dan daya komputasi yang memadai.

Meskipun sering kali diasosiasikan dengan risiko keamanan, teknologi *deepfake* di Indonesia mulai dimanfaatkan secara positif dalam industri kreatif dan preservasi sejarah, seperti upaya menghidupkan kembali narasi tokoh sejarah atau digunakan untuk kebutuhan *dubbing film* yang lebih presisi guna efisiensi produksi [21]. Namun, potensi manfaat tersebut dibayangi oleh sisi gelap yang mengancam ruang digital nasional melalui penyebaran disinformasi. Salah satu contoh kasus negatif yang sangat menggemparkan publik adalah beredarnya video palsu Presiden RI ke-7 Joko Widodo yang seolah-olah berpidato menggunakan bahasa Mandarin dengan fasih pada akhir tahun 2023 sebuah rekayasa yang bertujuan memicu sentimen politik dan keraguan publik terhadap integritas pemimpin negara [22].

Selain manipulasi politik, teknologi ini juga mulai merambah pada modus penipuan ekonomi dan pelanggaran privasi yang sangat merugikan individu. Di Indonesia, ditemukan kasus penyalahgunaan wajah tokoh publik, seperti Presiden Prabowo Subianto, dalam video *deepfake* yang digunakan untuk menawarkan bantuan sosial palsu demi mengelabui masyarakat agar menyerahkan data pribadi atau uang [23]. Fakta-fakta ini menegaskan bahwa tanpa literasi digital yang kuat dan pengawasan ketat, *deepfake* bukan sekadar inovasi teknologi, melainkan senjata siber yang mampu merusak tatanan sosial, menghancurkan reputasi seseorang melalui konten pornografi non-konsensual, hingga mengikis kepercayaan masyarakat terhadap kebenaran informasi visual [23].



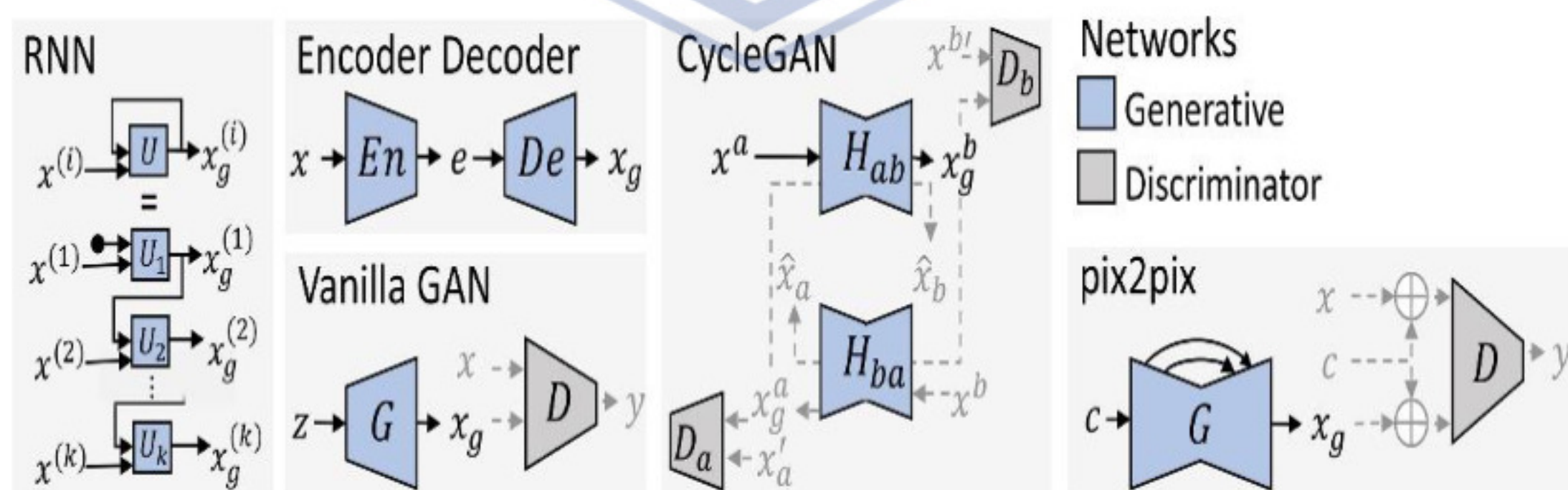
Gambar 2. 2 Contoh *Deepfake* [19].

2.2.2 Metode Pembuatan

Seperti yang terdapat pada Gambar 2.2 yaitu contoh sebuah pembuatan proses penciptaan *deepfake* secara fundamental yang bergantung pada arsitektur pembelajaran mendalam yang mampu memetakan fitur wajah dari satu individu ke individu lainnya. Berdasarkan tinjauan literatur, metode yang paling umum digunakan melibatkan struktur *Encoder-Decoder* (ED), di mana sebuah jaringan *encoder* bertugas mengekstraksi representasi wajah ke dalam ruang laten, sementara *decoder* bertugas membangun kembali citra tersebut agar sesuai dengan target. Jika dua *decoder* yang berbeda dilatih pada subjek yang berbeda namun menggunakan satu *encoder* yang sama, sistem dapat melakukan pertukaran wajah dengan cara memasukkan fitur wajah subjek A ke dalam *decoder* subjek B [19].

Selain model *autoencoder*, metode yang memberikan hasil paling realistis saat ini adalah *Generative Adversarial Networks* (GANs). Dalam mekanisme ini, dua model kecerdasan buatan saling berkompetisi: *generator* yang menciptakan citra sintetis, dan *discriminator* yang bertugas membedakan antara citra asli dan palsu. Persaingan ini memaksa *generator* untuk terus meningkatkan kualitas visual hingga mencapai titik di mana *discriminator* tidak dapat lagi mendeteksi rekayasa tersebut, menghasilkan konten media yang tampak sangat meyakinkan bagi mata manusia [19].

Secara lebih spesifik, pembuatan *deepfake* dapat diklasifikasikan ke dalam dua teknik utama, yaitu *replacement* (penggantian) dan *reenactment* (pemeranan ulang). Teknik *face swap* adalah contoh dari *replacement* di mana seluruh identitas wajah diganti, sedangkan *reenactment* melibatkan manipulasi ekspresi, gerakan mulut, atau posisi kepala dari target asli agar mengikuti instruksi dari sumber lain. Kedua teknik ini sering kali dipadukan dengan proses *post-processing* untuk menghaluskan batasan antara area yang direkayasa dengan latar belakang asli guna meminimalisir anomali visual [19].



Gambar 2. 3 Lima arsitektur jaringan saraf dasar untuk membuat *deepfake* [19].

Penjelasan pada Gambar 2.3:

- a) Representasi Arsitektur *Encoder-Decoder*: Gambar menunjukkan proses di mana wajah subjek (*Source*) dimasukkan ke dalam *Encoder* untuk diekstraksi fiturnya, kemudian diproses oleh *Decoder* untuk dipetakan ke wajah target. Ini sesuai dengan penjelasan paragraf pertama mengenai manipulasi ruang laten.
- b) *Visualisasi Face Swap (Replacement)*: Bagian bawah gambar dengan jelas memperlihatkan hasil akhir di mana identitas wajah dari *Source* (pria berbaju biru) berhasil ditempelkan ke kepala *Target* (pria berbaju merah). Ini adalah penggambaran sempurna untuk teknik *replacement* yang disebutkan di paragraf ketiga.
- c) Proses Pelatihan (*Training*): Alur panah yang menunjukkan transisi dari gambar asli ke gambar hasil rekonstruksi (*reconstructed image*) mewakili fase pembelajaran mesin di mana sistem berusaha meminimalkan kesalahan visual agar hasil rekayasa tampak identik dengan aslinya.

2.2.3 Deteksi Artefak Visual

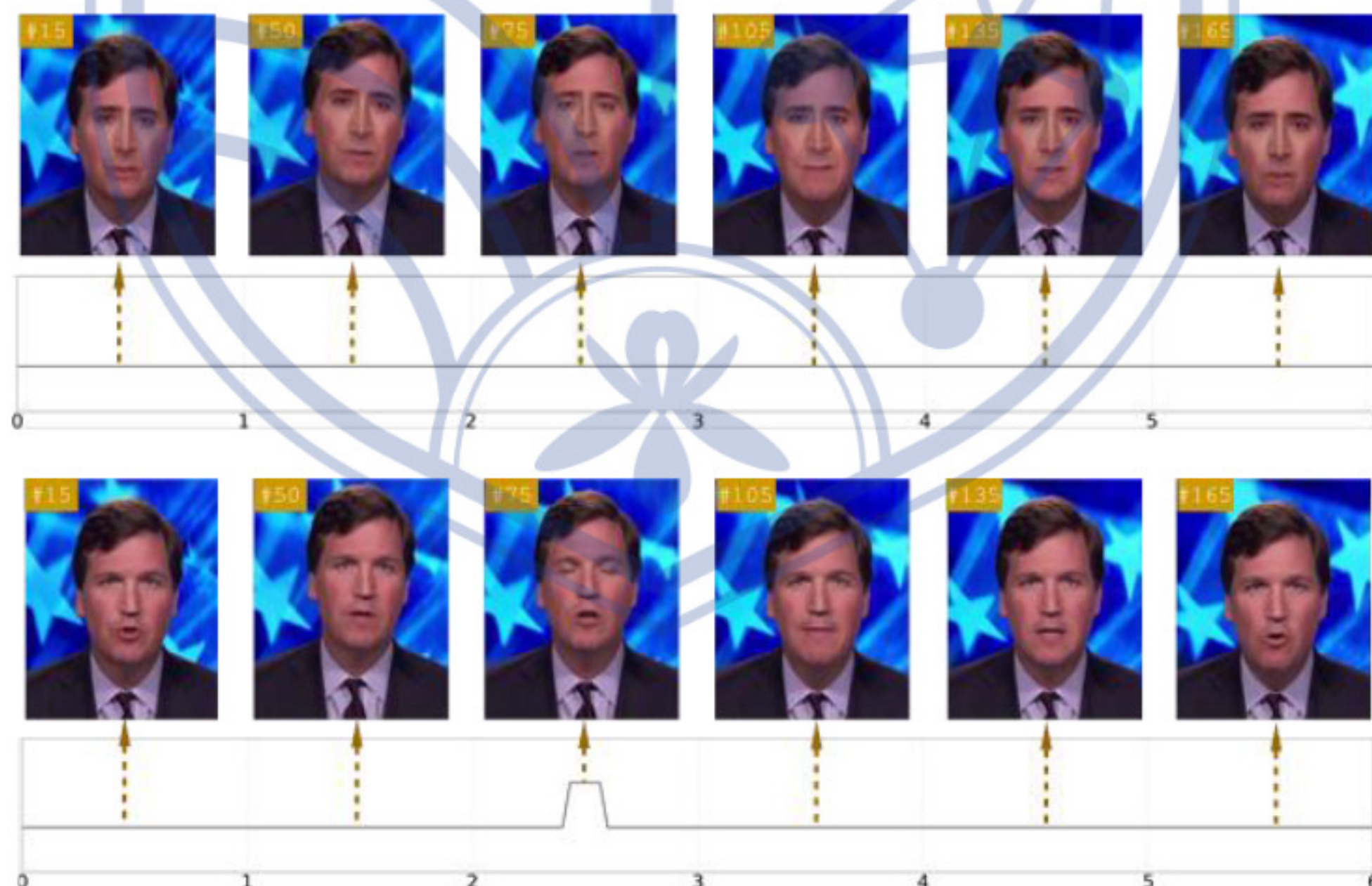
Meskipun teknologi manipulasi wajah semakin canggih, hasil rekayasa *deepfake* sering kali meninggalkan jejak digital yang tidak wajar atau dikenal sebagai artefak visual. Secara teknis, artefak ini muncul karena ketidakmampuan algoritma dalam mereplikasi detail biologis manusia secara sempurna, seperti pola kedipan mata yang tidak teratur atau gerakan bola mata yang tampak statis. Peneliti menemukan bahwa pada video palsu, subjek cenderung jarang berkedip atau memiliki ritme kedipan yang tidak sesuai dengan aktivitas fisiologis normal, karena data latih (gambar statis) sering kali tidak memberikan informasi yang cukup mengenai proses transisi kelopak mata manusia [24].

Selain anomali pada mata, deteksi artefak visual juga berfokus pada ketidakkonsistenan tekstur dan pencahayaan di sekitar batas wajah yang direkayasa. Di Indonesia, kajian mengenai keamanan siber menyoroti bahwa sering terjadi distorsi pada area tepi wajah, seperti bayangan yang tidak mengikuti arah sumber cahaya asli atau adanya efek "halus" yang berlebihan (*blurring*) pada kulit guna menyembunyikan sambungan antara wajah palsu dan latar belakang. Ketidakselarasan antara orientasi kepala dengan fitur wajah, misalnya posisi hidung yang sedikit melenceng dari sumbu wajah saat menoleh menjadi indikator kuat bahwa konten tersebut adalah hasil sintetis AI yang belum sempurna dalam memetakan geometri tiga dimensi [21].

Lebih jauh lagi, artefak visual dapat ditemukan melalui analisis spektral pada tingkat piksel yang tidak tertangkap oleh mata telanjang. Algoritma deteksi modern mampu mengidentifikasi sisa-sisa "sidik jari" digital yang ditinggalkan oleh jaringan saraf tiruan selama proses rekonstruksi gambar, seperti pola *noise* yang terlalu seragam atau artefak kompresi yang aneh pada area mulut saat berbicara [19]. Penemuan ini menunjukkan bahwa meskipun *deepfake* tampak sangat meyakinkan, integrasi antara pemrosesan data luar negeri dan pemantauan lokal di Indonesia tetap mengandalkan identifikasi anomali fisik dan teknis ini sebagai garda terdepan dalam memvalidasi keaslian sebuah informasi visual [21].

2.3 Pengolahan Citra Digital

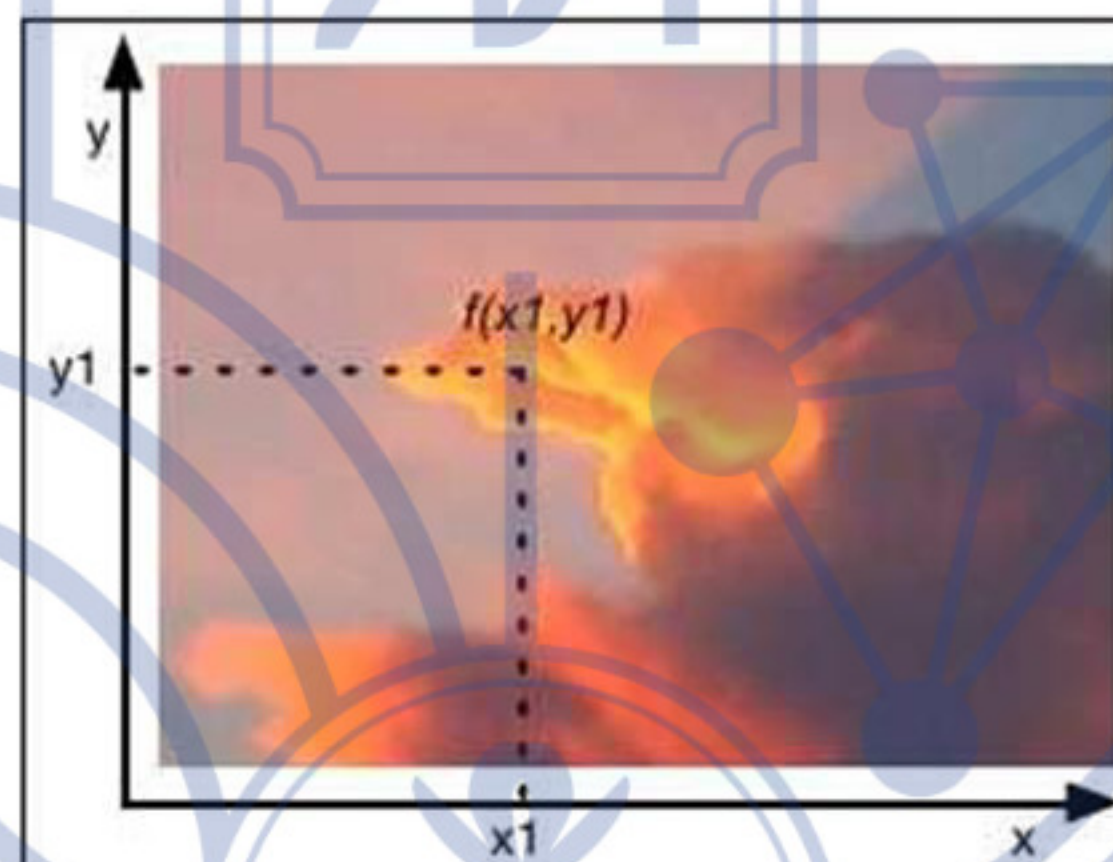
Pengolahan Citra Digital merupakan suatu disiplin ilmu yang berfokus pada manipulasi gambar dua dimensi menggunakan sistem komputer untuk memperbaiki kualitas visual atau mengekstraksi informasi tertentu. Secara teknis, citra digital dipahami sebagai kumpulan elemen terkecil yang disebut piksel, di mana setiap piksel memiliki nilai numerik yang merepresentasikan tingkat kecerahan atau warna dalam bentuk matriks. Proses ini sangat krusial karena memungkinkan data visual mentah diubah menjadi format yang lebih informatif dan mudah dianalisis, baik oleh mata manusia maupun oleh algoritma cerdas [25]. Contohnya seperti di Gambar 2.4 berikut ini.



Gambar 2. 4 Contoh deteksi kedipan mata pada video asli dan video *deepfake* [24].

Dalam praktiknya, pengolahan citra sering kali melibatkan operasi matematika kompleks untuk mengatasi berbagai kendala pada gambar asli, seperti kekaburan (*blur*) atau gangguan visual (*noise*). Salah satu teknik yang paling umum digunakan adalah metode konvolusi, yang bekerja dengan cara menggeser matriks kecil (*kernel*) di atas citra untuk memberikan efek tertentu [26]. Melalui teknik ini, sebuah gambar dapat diproses secara otomatis untuk menghasilkan efek visual yang lebih tajam, halus, atau bahkan menonjolkan tekstur tertentu secara lebih efektif dibandingkan dengan pengolahan manual yang memakan waktu lama.

Lebih jauh lagi, aplikasi dari pengolahan citra digital mencakup berbagai spektrum kebutuhan, mulai dari kebutuhan estetika di studio fotografi hingga analisis teknis untuk deteksi tepi objek. Dengan menerapkan variasi pengolahan seperti *sharpening*, *Gaussian blur*, hingga *edge detection*, sistem dapat membedakan batas-batas objek dalam citra dengan tingkat presisi yang tinggi [26]. Pemahaman mendalam mengenai struktur piksel dan algoritma pemrosesan seperti di Gambar 2.5 ini menjadi fondasi utama dalam pengembangan teknologi visi komputer yang lebih maju di masa depan.



Gambar 2. 5 Citra Digital [26].

2.3.1 Konsep Dasar dan Resolusi Citra

Konsep dasar pengolahan citra digital berakar pada representasi gambar sebagai fungsi intensitas cahaya dalam bidang dua dimensi. Secara matematis, citra ini dimodelkan sebagai matriks $f(x, y)$, di mana x dan y merupakan koordinat spasial, sementara nilai pada titik tersebut melambangkan tingkat kecerahan atau warna yang disebut sebagai intensitas [27]. Agar sebuah gambar dapat diproses oleh perangkat komputer, data visual yang bersifat kontinu harus dikonversi menjadi bentuk numerik atau nilai diskrit, sehingga sistem dapat mengenali dan memanipulasi setiap elemen penyusunnya secara akurat [27].

Elemen terkecil yang menyusun matriks citra tersebut dikenal dengan istilah piksel (*picture element*). Resolusi sebuah citra sangat ditentukan oleh kerapatan piksel-piksel ini; semakin banyak jumlah piksel yang tersusun dalam baris dan kolom, maka detail informasi yang terkandung dalam gambar akan semakin tajam [25]. Selain dimensi spasial, terdapat juga konsep resolusi kecerahan yang menentukan seberapa halus gradasi warna yang dapat ditampilkan, yang biasanya dinyatakan dalam bit per piksel untuk merepresentasikan tingkat keabuan atau intensitas warna tertentu.

Dalam praktiknya, pemahaman terhadap resolusi dan struktur piksel ini menjadi kunci utama sebelum melakukan transformasi citra yang lebih kompleks. Misalnya, saat menerapkan metode konvolusi untuk menajamkan atau menghaluskan gambar, sistem akan bekerja dengan mengolah nilai-nilai pada setiap piksel berdasarkan matriks tetangganya melalui kernel tertentu [26]. Dengan mengatur parameter resolusi dan pengolahan piksel yang tepat, kualitas citra dapat ditingkatkan untuk berbagai kebutuhan, mulai dari deteksi objek otomatis hingga keperluan estetika pada industri fotografi digital.

2.3.2 Ruang Warna RGB

Ruang warna RGB (*Red, Green, Blue*) merupakan model warna yang paling dasar dan paling luas digunakan dalam sistem pengolahan citra digital maupun perangkat elektronik. Konsep utama dari model ini adalah sistem penambahan atau *additive color model*, di mana berbagai variasi warna diciptakan dengan mengombinasikan tiga warna primer, yaitu merah, hijau, dan biru, dengan intensitas cahaya yang berbeda-beda. Secara matematis, model RGB dapat divisualisasikan sebagai sebuah ruang koordinat kartesius tiga dimensi atau berbentuk kubus, di mana setiap sumbu mewakili komponen warna primer tersebut. Titik asal (0,0,0) merepresentasikan warna hitam (tidak ada cahaya), sedangkan nilai maksimum pada semua sumbu akan menghasilkan warna putih murni [27].

Dalam struktur citra digital, setiap piksel pada model RGB menyimpan informasi warna dalam bentuk kanal-kanal yang terpisah. Misalnya, dalam citra 24-bit, masing-masing kanal warna (merah, hijau, dan biru) dialokasikan sebesar 8 bit, yang berarti setiap warna primer memiliki rentang intensitas dari 0 hingga 255. Dengan demikian, kombinasi dari 3 warna ini mampu menghasilkan lebih dari 16 juta warna yang berbeda, yang memungkinkan komputer untuk menampilkan gambar dengan gradasi yang sangat halus dan detail yang menyerupai aslinya [25]. Ketajaman dan kualitas warna ini sangat bergantung pada resolusi citra, di mana kerapatan piksel yang tinggi akan memastikan representasi warna yang lebih akurat dan bebas dari efek pecah atau *aliasing* saat gambar diperbesar [25].

Penerapan model warna RGB sangat krusial dalam algoritma deteksi objek maupun rekayasa citra tingkat lanjut seperti *deepfake*. Dalam proses deteksi, sering kali dilakukan normalisasi warna pada model RGB untuk mengurangi pengaruh intensitas cahaya eksternal, sehingga sistem dapat mengenali objek berdasarkan karakteristik warna kulit atau fitur wajah dengan lebih konsisten [27]. Namun, dalam manipulasi wajah menggunakan metode seperti *Variational Autoencoder* (VAE), model RGB pada Gambar 2.6 dibawah ini menjadi dasar bagi jaringan saraf tiruan untuk mempelajari dan merekonstruksi tekstur wajah target. Proses ini melibatkan ekstraksi fitur warna dari data latih yang kemudian dipetakan kembali untuk menciptakan citra sintetis yang memiliki kemiripan visual yang sangat tinggi dengan wajah asli, namun tetap berbasis pada struktur matriks warna primer tersebut [28].



Gambar 2. 6 Contoh foto RGB [27].

Integrasi antara pemahaman ruang warna dan teknik manipulasi digital ini memungkinkan pengembangan sistem forensik yang mampu mengidentifikasi jejak rekayasa pada tingkat piksel. Analisis terhadap distribusi nilai pada kanal merah, hijau, dan biru sering kali menyingkap ketidakkonsistenan spektral yang ditinggalkan oleh algoritma generator, terutama pada area transisi wajah [28]. Dengan menggabungkan metode konvolusi untuk memperjelas fitur dan analisis ruang warna yang mendalam, validasi terhadap keaslian sebuah konten media dapat dilakukan secara lebih akurat guna menghadapi ancaman disinformasi visual yang semakin canggih. Berdasarkan tinjauan referensi tersebut, berikut adalah beberapa data poin teknis yang penting.

a. Representasi Matriks: Citra digital dimodelkan sebagai matriks dua dimensi $f(x, y)$, di

mana setiap titik memiliki nilai intensitas diskrit [27].

- b. Kapasitas Warna: Model RGB 24-bit mampu menghasilkan total kombinasi sebanyak 16.581.375 variasi warna yang berbeda [25].
- c. Metode Konvolusi: Pengolahan nilai piksel dilakukan dengan menggeser matriks kernel (seperti 3x3) untuk menghasilkan efek penajaman atau deteksi tepi [25].
- d. Normalisasi Warna: Digunakan rumus $r = R / (R + G + B)$ untuk mendapatkan nilai warna yang stabil meski terjadi perubahan intensitas cahaya [27].
- e. Hasil Rekonstruksi VAE: Citra baru yang dihasilkan dari proses *training* VAE disebut sebagai gambar *decoder* atau hasil manipulasi [28].

2.3.3 Multi-task Cascaded Convolutional Networks (MTCNN)

Dalam konteks deteksi *deepfake*, tahapan pra-proses (*preprocessing*) yang akurat sangat krusial karena manipulasi umumnya berfokus pada area wajah. Untuk mengisolasi area tersebut dari keseluruhan citra digital, digunakan *Multi-task Cascaded Convolutional Networks* (MTCNN), sebuah *framework* berbasis jaringan saraf konvolusional mendalam yang mengintegrasikan tugas lokalisasi wajah (*face detection*) dan penentuan titik marka wajah (*face alignment*) [29]. Algoritma ini bekerja secara berjenjang (*cascaded*) melalui tiga tahapan jaringan konvolusi sekuensial setelah citra masukan diubah menjadi piramida citra (*image pyramid*), yaitu *Proposal Network* (P-Net) untuk menghasilkan kandidat kotak pembatas (*bounding box*) awal, *Refinement Network* (R-Net) untuk menyaring dan mengeliminasi prediksi palsu (*false positives*), serta *Output Network* (O-Net) untuk menentukan koordinat akhir wajah secara presisi sekaligus mengekstrak lima titik marka kunci wajah yang meliputi posisi mata, hidung, dan ujung mulut [29].

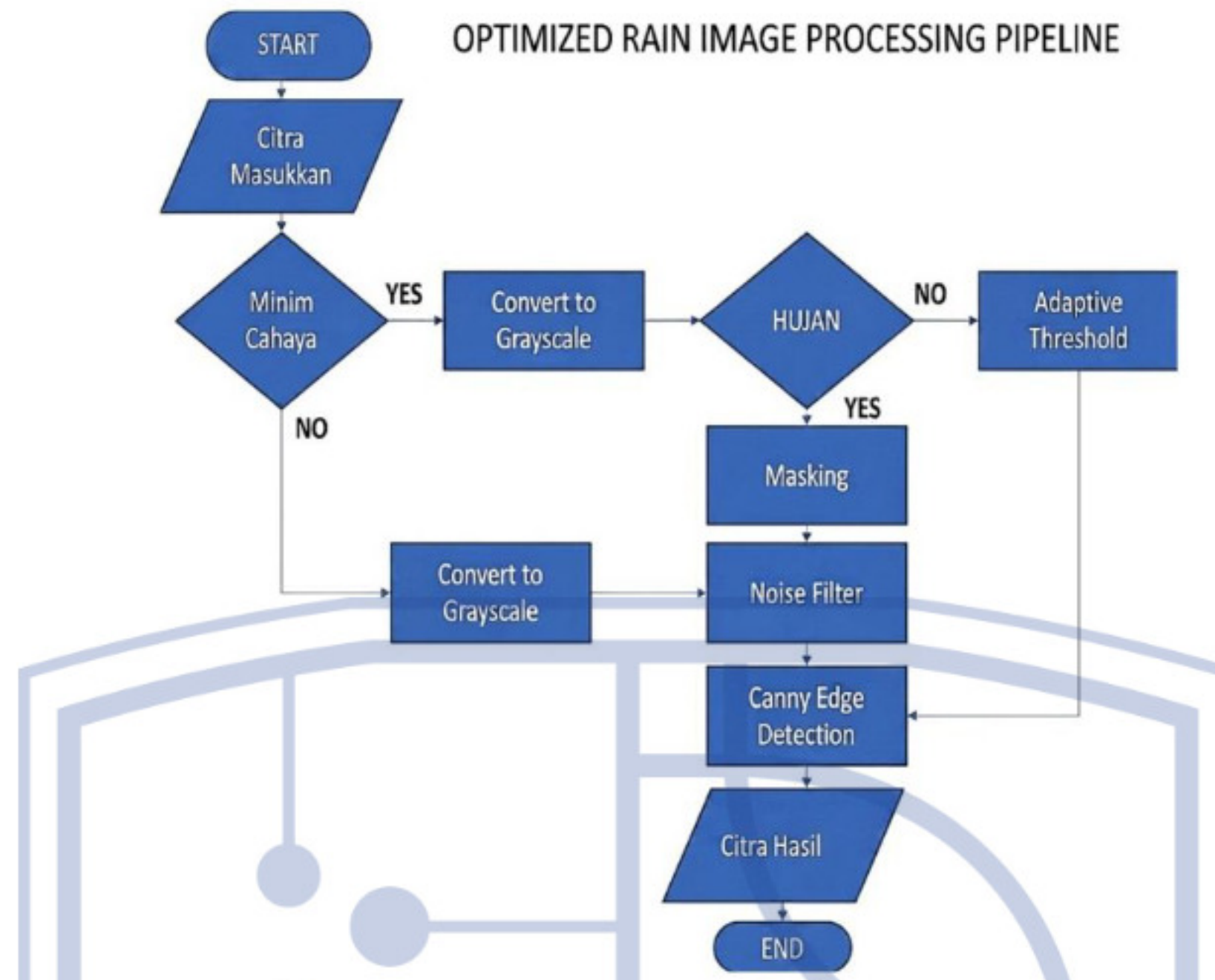
Penggunaan MTCNN dalam penelitian ini memberikan kontribusi signifikan terhadap performa model *Convolutional Neural Network* (CNN) utama yang bertindak sebagai klasifikator *deepfake*. Melalui informasi dari lima titik marka wajah yang dihasilkan oleh O-Net, citra wajah yang memiliki variasi posisi atau kemiringan dapat ditransformasikan secara konsisten ke posisi tegak (*crop dan align*) [30]. Proses standarisasi ini berhasil mereduksi gangguan latar belakang (*background noise*) yang tidak relevan, sehingga model CNN utama dapat memusatkan proses ekstraksi fitur secara optimal pada area wajah untuk mendeteksi anomali mikro seperti ketidakkonsistenan tekstur kulit, distorsi sekitar mata, atau artefak digital yang mengindikasikan adanya manipulasi *deepfake* demi pencegahan penyebaran hoaks visual di media sosial [29].

2.3.4 Prapemrosesan Citra (*Image Preprocessing*)

Prapemrosesan citra (*image preprocessing*) merupakan tahapan awal dalam pengolahan citra digital yang bertujuan untuk memperbaiki kualitas citra sebelum dilakukan analisis lebih lanjut. Citra digital pada dasarnya direpresentasikan sebagai kumpulan piksel dalam bentuk matriks dua dimensi, sehingga diperlukan proses awal agar data tersebut lebih siap digunakan dalam sistem komputasi [7]. Tahap ini menjadi krusial karena kualitas citra masukan sangat memengaruhi akurasi proses berikutnya, seperti deteksi, klasifikasi, maupun segmentasi objek.

Tahapan prapemrosesan umumnya meliputi beberapa teknik seperti normalisasi, reduksi noise, peningkatan kontras, serta transformasi ruang warna. Salah satu teknik yang sering digunakan adalah konversi citra dari format RGB menjadi *grayscale* untuk menyederhanakan kompleksitas data tanpa menghilangkan informasi penting. Selain itu, metode *filtering* seperti median filter atau *gaussian blur* digunakan untuk mengurangi gangguan *noise* sehingga fitur utama pada citra menjadi lebih jelas dan mudah dikenali oleh sistem [30]. Dalam penerapan pada sistem berbasis kecerdasan buatan, seperti model *Convolutional Neural Network* (CNN), prapemrosesan citra memiliki peran yang sangat penting. Tahapan ini mencakup proses penyesuaian ukuran citra, normalisasi data, serta peningkatan kualitas gambar agar *dataset* yang digunakan menjadi lebih konsisten [27]. Dengan data yang telah diproses dengan baik, model dapat mengekstraksi fitur dengan lebih optimal sehingga meningkatkan performa dalam proses klasifikasi atau pengenalan objek.

Sebagai ilustrasi, sebuah citra mentah (*raw image*) yang memiliki pencahayaan tidak merata dan banyak *noise* akan sulit dianalisis secara langsung. Namun, setelah melalui tahapan prapemrosesan seperti konversi *grayscale*, *filtering*, dan normalisasi, citra tersebut akan memiliki kualitas yang lebih baik dengan fitur yang lebih menonjol. Contohnya, data citra awal berupa matriks RGB (3 *channel*) dapat diubah menjadi matriks *grayscale* (1 *channel*) dengan intensitas yang lebih stabil, sehingga memudahkan proses deteksi pola. Hal ini menunjukkan bahwa prapemrosesan citra merupakan fondasi penting dalam keseluruhan proses pengolahan citra digital [27]. Berikut contoh kasus struktur berupa *Flowchart* dari Prapemrosesan Citra pada Gambar 2.7 beserta penjelasannya.



Gambar 2. 7 *Flowchart* Sistem kerja Penelitian[30].

- a. Tahap Masukan dan Konversi Citra: Proses diawali dengan Citra Masukkan, yaitu gambar mentah jalan raya yang ditangkap oleh kamera. Karena citra asli biasanya dalam format RGB (warna) yang memiliki beban komputasi tinggi, dilakukan tahap *Convert to Grayscale*. Langkah ini mengubah citra berwarna menjadi derajat keabuan untuk menyederhanakan data dengan tetap mempertahankan informasi intensitas cahaya yang diperlukan untuk mendeteksi garis marka [30].
- b. Segmentasi dan Fokus Area: Setelah citra menjadi abu-abu, dilakukan *Adaptive Thresholding*. Teknik ini memisahkan objek (marka jalan) dari latar belakang (aspal) berdasarkan perbedaan kontras lokal, sehingga marka jalan terlihat lebih menonjol meski kondisi pencahayaan jalan tidak merata. Selanjutnya, tahap *Masking* diterapkan untuk menentukan *Region of Interest* (ROI). *Masking* membatasi area pemrosesan hanya pada bagian bawah citra (area permukaan jalan) dan membuang bagian yang tidak relevan seperti langit atau pepohonan di pinggir jalan [30].
- c. Pembersihan dan Deteksi Tepi: Untuk memastikan hasil yang akurat, citra melewati *Noise Filter* (seperti *Gaussian* atau *Median Blur*) guna menghilangkan bintik-bintik kecil atau gangguan tekstur aspal yang bisa mengacaukan pendeteksian. Setelah citra bersih, dilakukan Deteksi Tepi (*Canny*). Algoritma *Canny* berfungsi untuk menemukan garis tepi yang tegas dari marka jalan dengan mencari perubahan intensitas piksel yang drastis [30].

- d. Interpretasi Garis dan Output: Tahap akhir adalah *Transformasi Hough*, sebuah metode matematika yang digunakan untuk menghubungkan titik-titik hasil deteksi tepi menjadi satu garis lurus yang utuh. Setelah garis marka teridentifikasi secara visual, sistem melakukan Penentuan Koordinat untuk mendapatkan nilai numerik posisi marka tersebut di dalam ruang dua dimensi. Data koordinat inilah yang menjadi output akhir sebagai representasi posisi marka jalan yang terbaca oleh sistem [30].

2.4 Kecerdasan Buatan (*Artificial Intelligence*)

Kecerdasan Buatan atau *Artificial Intelligence* (AI) merupakan inovasi teknologi yang dirancang untuk meniru kemampuan kognitif manusia, terutama dalam memecahkan masalah dan melakukan pengambilan keputusan secara otomatis. Teknologi ini bekerja dengan cara memproses data dalam jumlah besar melalui algoritma tertentu agar mesin dapat belajar dari pola yang ditemukan dan meningkatkan akurasi secara mandiri seiring waktu [31]. Dalam ekosistem digital saat ini, AI telah menjadi penggerak utama dalam transformasi digital yang menyentuh berbagai aspek kehidupan sosial dan ekonomi masyarakat modern [32]. Implementasi AI yang tepat terbukti mampu memberikan solusi cerdas terhadap tantangan kompleks yang sebelumnya sulit diselesaikan melalui metode konvensional atau pemrosesan manual.

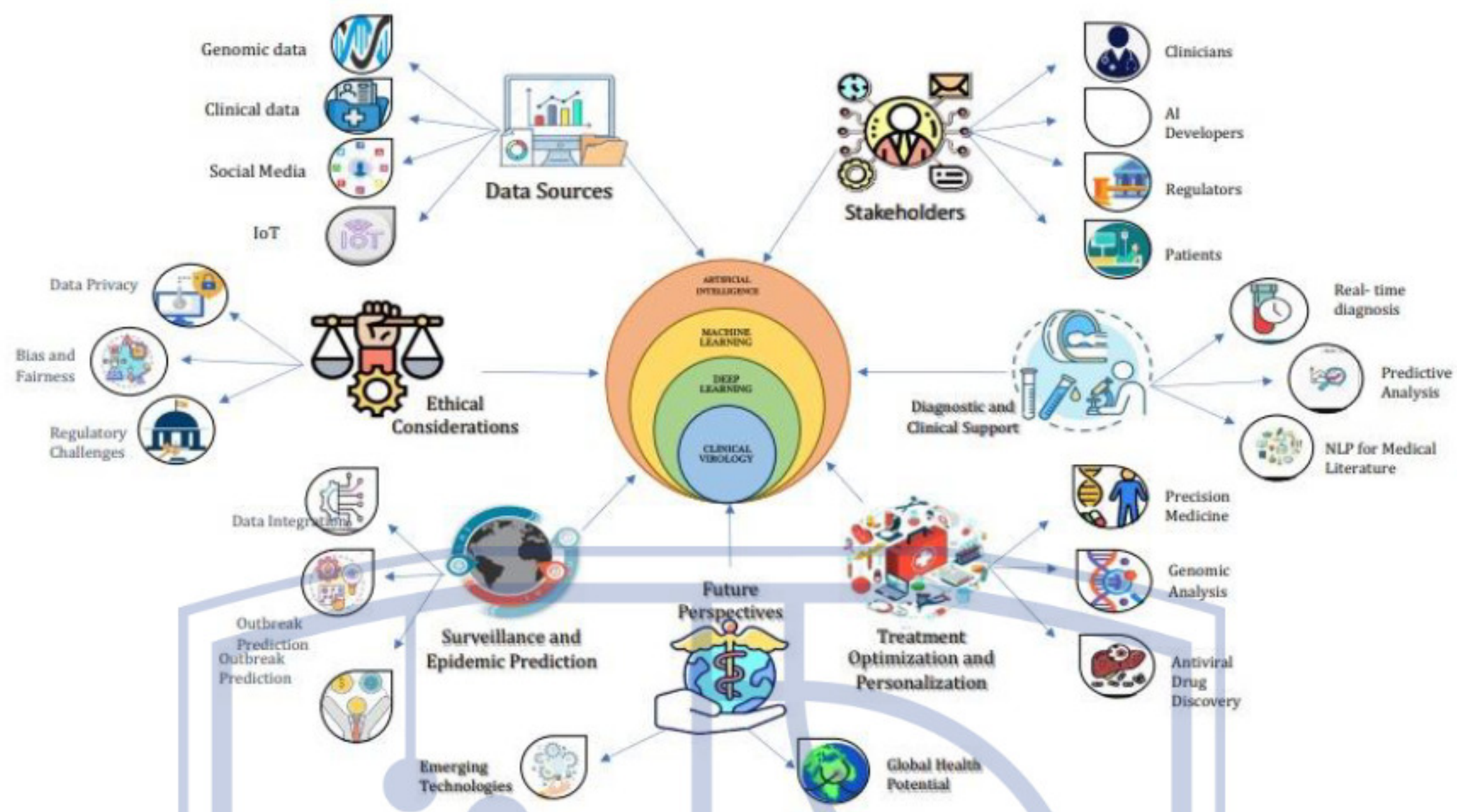
Dalam dunia bisnis dan industri, penerapan kecerdasan buatan fokus pada optimalisasi proses operasional guna mencapai efisiensi yang lebih tinggi dan meminimalisir kesalahan manusia. Sistem cerdas ini mampu melakukan analisis prediktif terhadap perilaku pasar, mengelola manajemen rantai pasok secara otomatis, hingga menyediakan layanan pelanggan melalui asisten virtual yang responsif. Dengan mengintegrasikan AI ke dalam struktur organisasi, perusahaan dapat mempercepat alur kerja dan mengurangi biaya produksi melalui pemanfaatan mesin yang mampu bekerja tanpa henti dengan standar kualitas yang konsisten. Hal ini menjadikan AI sebagai instrumen strategis yang memungkinkan entitas bisnis untuk tetap kompetitif di tengah persaingan pasar global yang semakin dinamis [31].

Dalam hierarki teknologi ini, Kecerdasan Buatan menaungi sub-bidang yang lebih spesifik, yaitu *Machine Learning* dan *Deep Learning*, yang menjadi mesin penggerak utama di balik kemampuan adaptif sistem cerdas. *Machine Learning* memungkinkan AI untuk mempelajari pola dari dataset yang ada tanpa perlu diprogram secara eksplisit untuk setiap tugas, sementara *Deep Learning* menggunakan arsitektur jaringan saraf tiruan berlapis untuk memproses data visual yang sangat kompleks. Integrasi kedua teknologi ini memungkinkan

sistem melakukan analisis tingkat tinggi, mulai dari klasifikasi objek otomatis pada citra digital hingga pendeteksian anomali yang sangat halus yang tidak dapat ditangkap oleh metode statistik konvensional. Dengan memanfaatkan struktur pembelajaran mendalam ini, sistem AI mampu memahami hubungan antar fitur dalam sebuah dataset yang masif secara mandiri [31].

Namun, perkembangan ini juga membawa tantangan etis terkait manipulasi informasi visual melalui teknik rekayasa wajah atau *deepfake*. Algoritma *deep learning* tingkat lanjut kini mampu menciptakan konten media sintetis yang sangat realistis, yang jika disalahgunakan, dapat digunakan untuk penipuan atau penyebaran disinformasi yang merugikan identitas seseorang. Fenomena ini memaksa para pengembang untuk menciptakan sistem pertahanan berbasis AI yang mampu mendeteksi artefak atau anomali pada tingkat piksel yang dihasilkan oleh mesin generator itu sendiri. Oleh karena itu, diperlukan kerangka regulasi yang kuat dan pengawasan hukum yang transparan untuk memastikan teknologi ini digunakan secara bertanggung jawab tanpa mengabaikan aspek keamanan digital [28].

Efektivitas sistem AI sangat bergantung pada kualitas data input dan ketepatan tahap prapemrosesan sebelum algoritma melakukan pembelajaran mendalam. Data visual yang telah melalui tahap normalisasi dan pembersihan memungkinkan jaringan saraf tiruan untuk mengenali pola dengan lebih akurat, baik itu untuk keperluan optimasi bisnis maupun untuk membedakan antara konten asli dan hasil rekayasa sintetis [31]. Sinergi antara inovasi teknologi, kesadaran etika, dan perlindungan hukum akan memastikan bahwa kecerdasan buatan tetap menjadi alat yang memberdayakan manusia dalam mencapai efisiensi maksimal tanpa mengorbankan keamanan data pribadi [32]. Berikut gambaran dari struktur konsep dari *Artificial Intelligence (AI)*, *Machine Learning*, dan *Deep Learning* yang ada pada Gambar 2.8 dibawah ini.



Gambar 2. 8 Kerangka kerja terintegrasi AI dalam virologi klinis dan epidemiologi [33].

Gambar tersebut mengilustrasikan hubungan inklusif antara Kecerdasan Buatan (AI), *Machine Learning*, dan *Deep Learning*, di mana masing-masing merupakan bagian dari entitas yang lebih besar. Kecerdasan Buatan (AI) menempati lapisan terluar sebagai konsep payung yang mencakup pengembangan sistem komputer untuk melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia. Dalam konteks penelitian ini, AI menjadi kekuatan utama yang mendorong perubahan transformatif dalam presisi diagnostik, intervensi terapeutik, hingga pemantauan epidemiologi secara komprehensif [33]. Lapisan kedua, yaitu *Machine Learning*, dijelaskan sebagai sub-bidang dari AI yang berfokus pada penggunaan algoritma untuk belajar secara otomatis dari data tanpa instruksi pemrograman eksplisit. *Machine Learning* memungkinkan komputer untuk mengenali pola-pola rumit dalam dataset virologi yang besar, seperti mengoptimalkan pengurutan genom virus dan mempercepat penemuan obat melalui wawasan prediktif. Fleksibilitas *Machine Learning* dalam memproses data historis menjadikannya alat yang sangat efektif untuk meramalkan wabah virus di masa depan berdasarkan tren yang terdeteksi [33]. Lapisan terdalam adalah *Deep Learning*, yang merupakan evolusi lebih lanjut dan sangat terspesialisasi dari *Machine Learning*. *Deep Learning* menggunakan arsitektur jaringan saraf tiruan berlapis-lapis untuk mengekstraksi fitur secara otomatis dari data yang sangat kompleks dan tidak terstruktur. Contoh nyata penerapannya adalah penggunaan algoritma *Deep Learning* untuk memproses gambar CT guna melakukan penyaringan (*screening*) penyakit virus seperti COVID-19 secara cepat dan akurat, yang melampaui kemampuan analisis statistik tradisional [33].

Secara keseluruhan, integrasi ketiga lapisan teknologi ini menciptakan sinergi yang memungkinkan para peneliti virologi dan ilmuwan data untuk berkolaborasi lebih erat. Meskipun membawa kemajuan besar, dokumen tersebut menekankan bahwa penerapan AI, *Machine Learning*, dan *Deep Learning* harus tetap memperhatikan tantangan terkait keamanan data, bias algoritma, dan pertimbangan etika. Dengan pengelolaan yang tepat, hierarki teknologi ini diproyeksikan akan terus menjadi fondasi utama dalam menghadapi evolusi penyakit virus yang cepat di masa mendatang [33].

2.4.1 *Machine Learning*

Machine Learning merupakan salah satu pilar utama dari kecerdasan buatan yang berfokus pada pengembangan sistem agar mampu belajar secara mandiri dari data tanpa harus diprogram secara eksplisit untuk setiap tugas. Teknologi ini bekerja dengan menggunakan algoritma statistik untuk mengidentifikasi pola-pola rumit dan korelasi tersembunyi di dalam dataset yang sangat besar dan kompleks [34]. Dengan memproses informasi secara iteratif, sistem *machine learning* dapat terus meningkatkan performanya seiring dengan bertambahnya jumlah data yang dipelajari. Hal ini memungkinkan transisi dari pemrosesan data manual menuju sistem otomatisasi cerdas yang lebih efisien dan akurat dalam berbagai disiplin ilmu.

Secara teknis, *Machine Learning* dapat diklasifikasikan menjadi tiga kategori utama, yaitu pembelajaran terarah (*supervised learning*), pembelajaran tidak terarah (*unsupervised learning*), dan pembelajaran penguatan (*reinforcement learning*). Pembelajaran terarah menggunakan data yang telah diberi label untuk melatih model, sementara pembelajaran tidak terarah mencari struktur alami atau pengelompokan dalam data mentah yang tidak berlabel. Namun, *reinforcement learning* memungkinkan agen cerdas untuk belajar melalui interaksi dengan lingkungan guna mendapatkan hasil maksimal berdasarkan sistem penghargaan. Ketiga kategori ini terus dikembangkan untuk mengurangi beban komputasi serta mempercepat kinerja sistem dalam mencapai tingkat presisi yang tinggi. Dalam implementasi praktisnya di bidang medis dan virologi, *machine learning* berperan krusial dalam mempercepat penemuan obat melalui analisis prediktif terhadap data sejarah yang masif. Algoritma *Machine Learning* mampu mengoptimalkan penjujukan genom virus serta memberikan wawasan berharga dalam peramalan wabah penyakit di masa depan berdasarkan tren epidemiologi yang terdeteksi. Lalu teknologi ini memfasilitasi kolaborasi antara ilmuwan data dan peneliti medis untuk menciptakan intervensi terapeutik yang lebih tepat sasaran. Pemanfaatan *machine learning* terbukti signifikan dalam meningkatkan

akurasi diagnostik serta efektivitas pemantauan kesehatan masyarakat secara luas [34].

Meskipun menawarkan kemajuan yang luar biasa, pengembangan *machine learning* menghadapi tantangan besar terkait keamanan data, potensi bias pada algoritma, serta pertimbangan etika dalam penggunaannya. Kualitas output dari model *machine learning* sangat bergantung pada integritas data yang digunakan selama fase pelatihan, sehingga tahap prapemrosesan data menjadi sangat vital untuk menghindari hasil yang menyesatkan. Dibutuhkan tata kelola yang transparan dan kolaborasi antardisiplin untuk memastikan bahwa inovasi *Machine Learning* tetap memberikan manfaat positif bagi masyarakat. Dengan pengelolaan yang tepat, *machine learning* diproyeksikan akan terus menjadi fondasi utama bagi pengembangan teknologi masa depan yang lebih adaptif dan responsif terhadap tantangan global [34]. Berikut alur kerja dari *machine learning*.



Gambar 2. 9 Arsitektur dan alur kerja *classic machine learning* [35].

Dalam alur kerjanya pada Gambar 2.9 seperti di atas, sistem ini melibatkan proses ekstraksi fitur secara manual di mana karakteristik tertentu dari data input dipilih untuk melatih model agar dapat melakukan klasifikasi atau prediksi. Berbeda dengan pendekatan tradisional, *machine learning* memanfaatkan teknik statistik untuk mengidentifikasi korelasi antar variabel, yang dalam konteks medis sangat berguna untuk membantu diagnosis penyakit berdasarkan analisis gambar radiologi secara otomatis [35].

Penerapan *Machine Learning* memerlukan tahapan prapemrosesan yang ketat untuk memastikan integritas data sebelum digunakan dalam fase pelatihan model. Efektivitas dari model *machine learning* ini sangat bergantung pada kualitas pemilihan fitur yang relevan, sehingga kolaborasi antara ilmuwan data dan pakar domain menjadi sangat krusial untuk menghasilkan *output* yang akurat. Meskipun metode ini telah membawa kemajuan besar dalam otomatisasi analisis data, tantangan seperti bias algoritma dan kebutuhan akan interpretasi model (*Explainable AI*) tetap menjadi fokus utama untuk memastikan teknologi ini aman dan transparan dalam penggunaan praktisnya [35].

2.4.2 Deep Learning

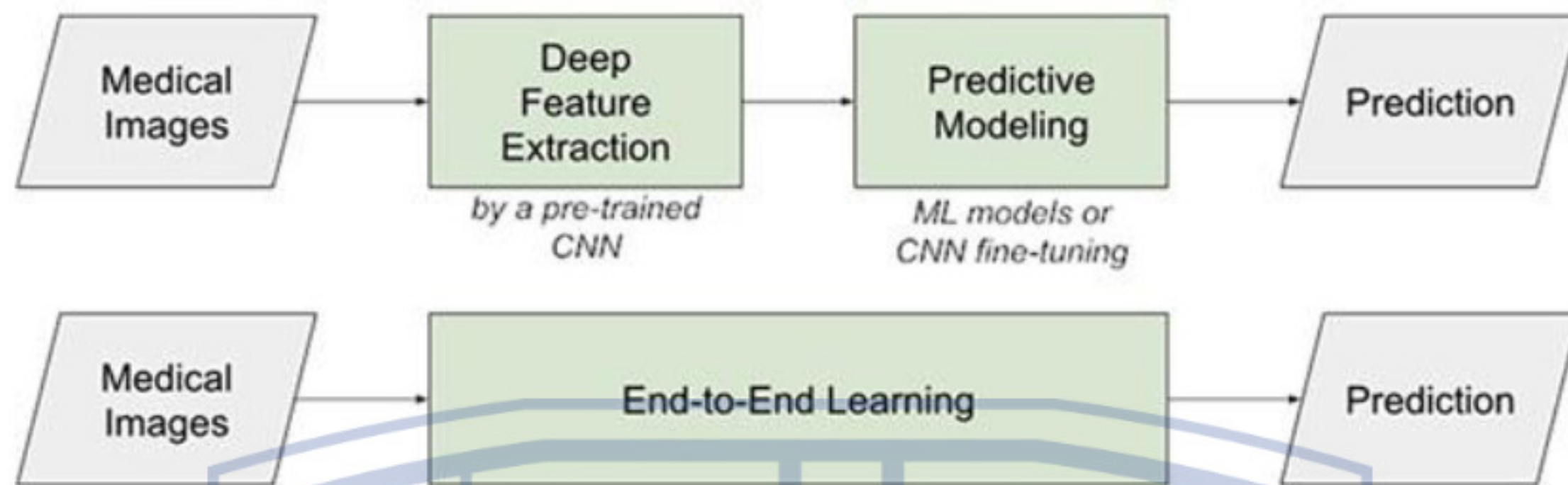
Deep learning merupakan evolusi mutakhir dari kecerdasan buatan yang menggunakan arsitektur jaringan saraf tiruan berlapis untuk memproses data dengan tingkat kompleksitas yang tinggi. Berbeda dengan metode konvensional, teknologi ini

memungkinkan mesin untuk secara otomatis mempelajari representasi data melalui hierarki konsep yang dibangun dari lapisan-lapisan saraf tiruan tersebut [36]. Kemampuannya dalam mengekstraksi fitur secara mandiri tanpa memerlukan intervensi manual menjadikan *deep learning* sebagai instrumen yang sangat efisien untuk mengolah *dataset* besar yang tidak terstruktur. Dengan meniru cara kerja otak manusia dalam mengenali pola, sistem ini mampu mencapai tingkat akurasi yang melampaui teknik pembelajaran mesin tradisional dalam berbagai domain aplikasi.

Dalam implementasi teknisnya, *deep learning* sangat bergantung pada ketersediaan data latih yang masif dan daya komputasi yang besar untuk melatih model-model jaringan saraf yang dalam. Struktur intinya terdiri dari lapisan input, beberapa lapisan tersembunyi (*hidden layers*), dan lapisan output yang bekerja secara sinergis untuk melakukan transformasi data dari bentuk mentah menjadi informasi yang bermakna [34]. Penggunaan algoritma seperti *Convolutional Neural Networks* (CNN) telah menjadi standar dalam pengolahan citra digital karena kemampuannya dalam mengidentifikasi detail visual yang sangat halus dan spesifik secara otomatis. Hal ini memungkinkan pengembangan sistem otomatisasi yang lebih cerdas dan responsif terhadap variasi input data yang dinamis di lapangan. Penerapan *deep learning* telah membawa transformasi signifikan, salah satunya dalam bidang medis di mana teknologi ini digunakan untuk menganalisis citra radiologi dengan presisi diagnostik yang sangat tinggi. Model *deep learning* mampu mendeteksi anomali pada gambar medis seperti hasil pemindaian CT atau MRI secara lebih cepat dan akurat dibandingkan dengan analisis manual oleh pakar [35]. Selain di bidang kesehatan, konsep ini juga mulai merambah dunia pendidikan melalui pendekatan pembelajaran yang lebih bermakna (*meaningful*), mendalam (*mindful*), dan menyenangkan (*joyful*) untuk meningkatkan pemahaman kognitif peserta didik. Fleksibilitas arsitekturnya menjadikan *deep learning* sebagai solusi serbaguna yang dapat diadaptasi untuk berbagai tantangan riset dan kebutuhan praktis masyarakat modern.

Meskipun menawarkan potensi yang luar biasa, penggunaan *deep learning* memerlukan pengawasan ketat terkait integritas data dan transparansi algoritma guna menghindari bias hasil. Proses pelatihan yang kompleks sering kali menjadikan model ini bersifat seperti "kotak hitam" yang sulit diinterpretasikan secara langsung proses pengambilan keputusannya, sehingga pengembangan metode *Explainable AI* (XAI) menjadi sangat relevan [34]. Selain itu, koordinasi antara pakar domain dan ilmuwan data sangat diperlukan untuk memastikan bahwa model yang dikembangkan benar-benar valid dan aman untuk diimplementasikan secara luas. Dengan pengelolaan dan etika yang tepat, *deep*

learning diproyeksikan akan terus menjadi teknologi kunci yang mendefinisikan masa depan otomatisasi cerdas dan analisis data global [35].



Gambar 2. 10 Arsitektur dan Alur kerja *Deep Learning* [35].

Berbeda dengan metode pembelajaran mesin tradisional yang memerlukan pemilihan fitur secara manual oleh pakar, arsitektur *deep learning* pada Gambar 2.10 seperti di atas yang terdiri dari jaringan saraf tiruan berlapis mampu mengidentifikasi karakteristik penting langsung dari data input mentah. Dalam alur kerja yang ditunjukkan oleh gambar, proses ini memungkinkan sistem untuk melakukan klasifikasi objek dengan tingkat presisi yang sangat tinggi, terutama pada data visual kompleks seperti citra radiologi dalam dunia medis [35].

Efektivitas *Deep Learning* sangat bergantung pada ketersediaan dataset berskala besar dan daya komputasi yang kuat untuk melatih lapisan-lapisan saraf yang dalam. Karena kemampuannya dalam mengenali pola-pola non-linear yang rumit, *deep learning* sering kali dianggap sebagai "kotak hitam" yang memerlukan metode interpretasi tambahan agar keputusan yang diambil oleh model dapat dipahami secara logis oleh manusia. Meskipun demikian, integrasi *deep learning* dalam analisis citra digital tetap menjadi standar emas baru karena kemampuannya dalam memberikan hasil diagnostik yang lebih cepat dan konsisten dibandingkan dengan analisis konvensional [35].

2.5 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu metode dalam *deep learning* yang banyak digunakan untuk pengolahan citra digital karena kemampuannya dalam mengenali pola visual secara otomatis. CNN bekerja dengan meniru sistem visual manusia, di mana informasi dari citra diproses secara bertahap mulai dari fitur sederhana hingga fitur kompleks. Dalam implementasinya, CNN mampu melakukan ekstraksi fitur tanpa perlu proses manual, sehingga lebih efisien dan menghasilkan akurasi yang tinggi dalam berbagai tugas seperti klasifikasi citra, deteksi objek, dan pengenalan pola [5].

Salah satu keunggulan CNN terletak pada konsep *weight sharing* dan *local connectivity*, yang memungkinkan pengurangan jumlah parameter dalam jaringan sehingga proses pelatihan menjadi lebih cepat dan efisien. Selain itu, CNN juga menggunakan fungsi aktivasi seperti ReLU untuk memberikan sifat non-linear pada model, sehingga mampu mengenali pola yang kompleks dalam data citra. Proses pelatihan CNN dilakukan dengan bantuan fungsi *loss* dan algoritma optimasi seperti ADAM (*Adaptive Moment Estimation*) atau SGD (*Stochastic Gradient Descent*) untuk meminimalkan kesalahan prediksi secara bertahap [9].

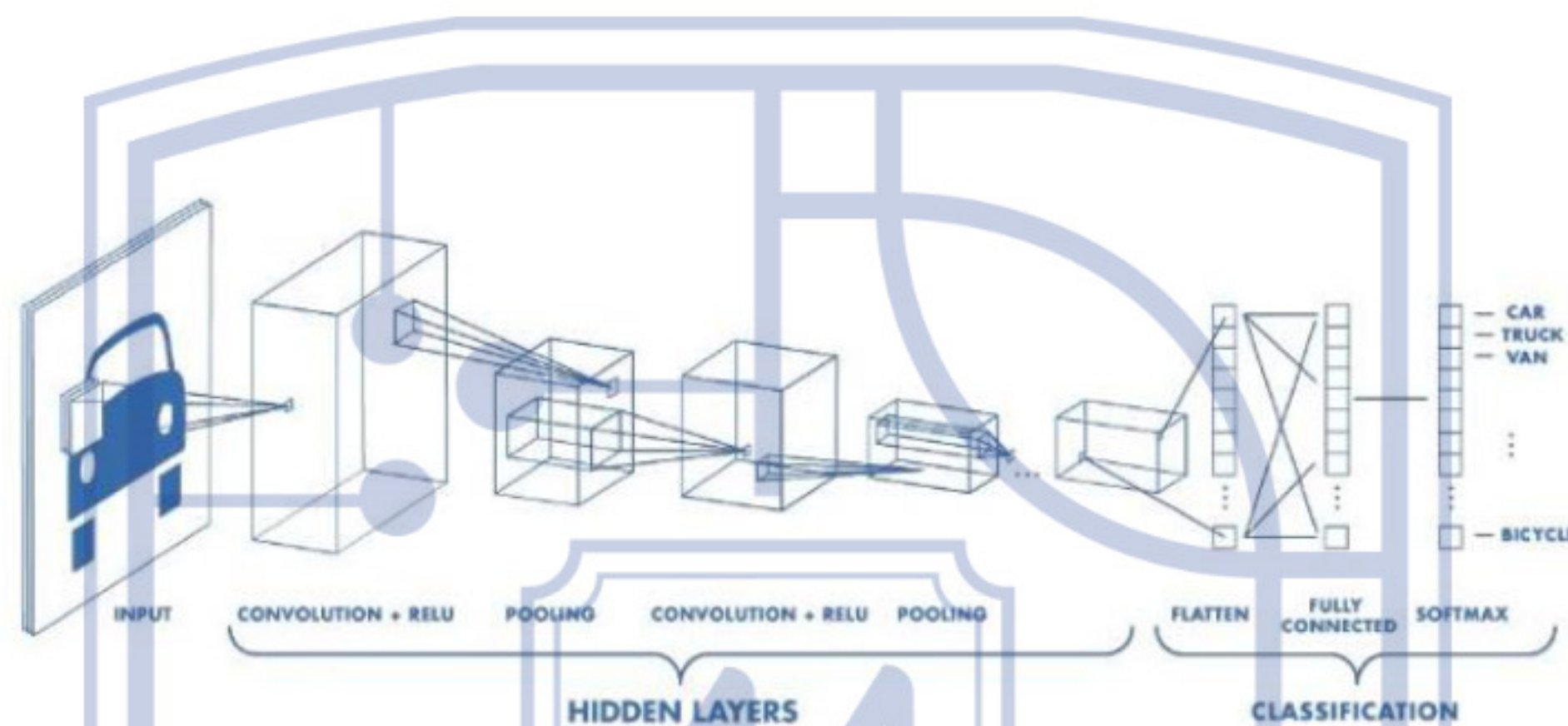
Dalam berbagai penelitian, CNN telah terbukti mampu memberikan performa yang tinggi dalam tugas klasifikasi citra, dengan tingkat akurasi yang baik tergantung pada kualitas *dataset* dan arsitektur yang digunakan. CNN banyak diterapkan dalam berbagai bidang seperti pengenalan wajah, klasifikasi objek, analisis citra medis, hingga deteksi konten manipulasi seperti *deepfake*. Dengan kemampuannya dalam mengekstraksi fitur visual secara otomatis dan mendalam, CNN menjadi metode yang sangat relevan untuk digunakan dalam penelitian deteksi citra manipulasi di media sosial [8].

2.5.1 Arsitektur Dasar CNN

Arsitektur dasar *Convolutional Neural Network* (CNN) dirancang untuk mengolah data citra dengan memanfaatkan struktur spasial yang terdapat pada gambar. Berbeda dengan jaringan saraf tradisional yang menggunakan input berupa vektor satu dimensi, CNN bekerja dengan data berdimensi tiga, yaitu tinggi (*height*), lebar (*width*), dan kedalaman (*depth*) yang merepresentasikan kanal warna pada citra. Struktur ini memungkinkan CNN untuk mempertahankan hubungan antar piksel sehingga proses ekstraksi fitur menjadi lebih optimal. CNN terdiri dari beberapa lapisan yang tersusun secara hierarkis, di mana setiap lapisan memiliki fungsi khusus dalam mengolah informasi visual secara bertahap [37].

Secara umum, arsitektur CNN terdiri dari beberapa komponen utama, yaitu lapisan konvolusi, fungsi aktivasi (*activation layer*), *pooling layer*, serta *fully connected layer*. Proses dimulai dari lapisan konvolusi yang berfungsi untuk mengekstraksi fitur dasar seperti tepi, sudut, dan tekstur dari citra. Hasil dari proses ini kemudian diteruskan ke fungsi aktivasi untuk memberikan sifat non-linear pada jaringan. Selanjutnya, *pooling layer* digunakan untuk mengurangi dimensi data sehingga mempercepat proses komputasi sekaligus mempertahankan informasi penting. Pada tahap akhir, fitur yang telah diekstraksi akan diproses oleh *fully connected layer* untuk menghasilkan output berupa klasifikasi atau prediksi [5].

Arsitektur CNN juga mengadopsi konsep *local connectivity* dan *weight sharing* yang menjadi keunggulan utama dibandingkan metode jaringan saraf konvensional. *Local connectivity* memungkinkan neuron hanya terhubung dengan sebagian kecil area dari input, sehingga fokus pada fitur lokal tertentu. Sementara itu, *weight sharing* memungkinkan penggunaan filter yang sama di seluruh bagian citra, sehingga jumlah parameter yang harus dipelajari menjadi lebih sedikit. Hal ini tidak hanya meningkatkan efisiensi komputasi, tetapi juga membantu model dalam mengenali pola yang sama pada berbagai posisi dalam citra [38].



Gambar 2. 11 Arsitektur *Convolutional Neural Network* (CNN) [5].

Adapun komponen utama dalam arsitektur CNN pada Gambar 2.11 meliputi:

1. *Convolution Layer*, merupakan lapisan utama yang berfungsi untuk mengekstraksi fitur dari citra dengan menggunakan kernel atau filter tertentu, sehingga mampu mendeteksi pola seperti tepi, tekstur, dan bentuk objek pada berbagai tingkat kompleksitas.
2. *Pooling Layer*, berfungsi untuk melakukan proses *downsampling* atau pengurangan dimensi fitur yang dihasilkan oleh lapisan konvolusi, sehingga dapat mengurangi kompleksitas komputasi serta membantu mengatasi *overfitting*.
3. *Fully Connected Layer*, merupakan lapisan akhir yang berfungsi untuk menghubungkan seluruh fitur yang telah diekstraksi menjadi sebuah keputusan klasifikasi berdasarkan hasil pembelajaran model [5].

Ketiga komponen tersebut bekerja secara terintegrasi dalam memproses citra dari tahap awal hingga menghasilkan output yang akurat, sehingga arsitektur CNN sangat efektif digunakan dalam berbagai aplikasi pengolahan citra, termasuk deteksi *deepfake* [5].

2.5.2 *Convolutional Layer*

Convolutional layer merupakan komponen utama dalam arsitektur *Convolutional Neural Network* (CNN) yang berfungsi untuk mengekstraksi fitur dari citra input melalui

operasi konvolusi. Proses ini dilakukan dengan menggeser sebuah kernel atau filter berukuran kecil pada seluruh bagian citra untuk menghasilkan *feature map* yang merepresentasikan pola tertentu seperti tepi, tekstur, maupun bentuk objek [39]. Setiap filter memiliki bobot yang akan dipelajari selama proses pelatihan, sehingga mampu mendeteksi karakteristik visual yang relevan secara otomatis. Selain itu, penggunaan beberapa filter dalam satu lapisan memungkinkan CNN untuk menangkap berbagai jenis fitur secara bersamaan [5].

Kernel atau filter dalam *convolutional layer* memiliki ukuran yang lebih kecil dibandingkan citra input, misalnya 3×3 atau 5×5 , dan berfungsi sebagai pendeteksi pola tertentu. Setiap kernel akan menghasilkan satu *feature map*, sehingga semakin banyak kernel yang digunakan, semakin banyak pula fitur yang dapat dipelajari oleh model. Proses ini memungkinkan CNN untuk mengenali berbagai karakteristik visual dalam citra secara otomatis tanpa perlu rekayasa fitur manual. Selain itu, penggunaan kernel yang sama di seluruh area citra (*weight sharing*) membuat proses komputasi menjadi lebih efisien dan mengurangi jumlah parameter yang harus dilatih [6]. Secara matematis, operasi konvolusi dapat dinyatakan sebagai berikut:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) \cdot K(m, n) / K = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \quad (1)$$

Keterangan:

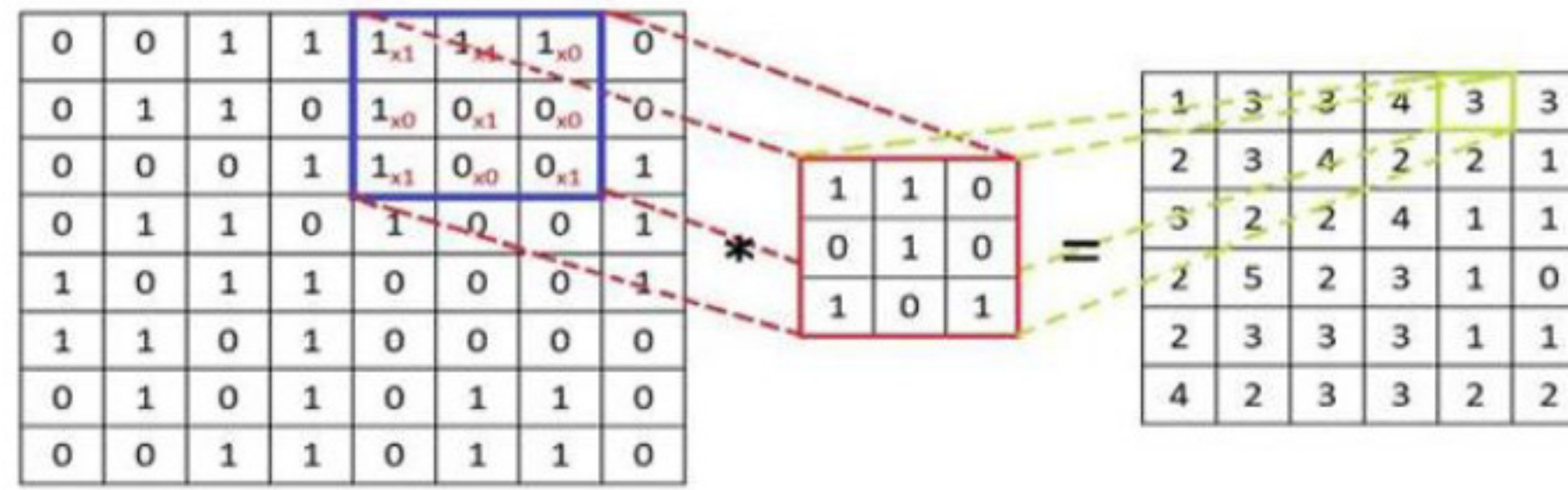
I = citra input

K = kernel atau filter

$S(i, j)$ = nilai output pada posisi (i, j) (*feature map*)

m, n = indeks ukuran kernel

Rumus tersebut menjelaskan bahwa setiap nilai pada *feature map* diperoleh dari hasil penjumlahan perkalian antara nilai piksel citra dan nilai kernel pada area tertentu. Proses ini dilakukan secara berulang di seluruh area citra, sehingga menghasilkan representasi fitur yang lebih sederhana namun *informative* [5].



Gambar 2. 12 *Convulutional Layer* [5].

Convulutional Layer yang terdapat pada Gambar 2.12 didefinisikan oleh: ukuran kernel, Panjang langkah, dan *padding*. Ukuran kernel adalah ukuran filter kernel atau kernel geser [22]. Panjang langkah adalah jumlah kernel yang bergeser sebelum membuat titik produk dan membuat piksel keluaran. *Padding* adalah ukuran bingkai ke-0 yang diatur di sekitar peta fitur masukan [5].

2.5.3 *Pooling Layer*

Pooling layer merupakan salah satu komponen penting dalam *Convolutional Neural Network* (CNN) yang berfungsi untuk melakukan reduksi dimensi terhadap *feature map* hasil dari *convolutional layer*. Proses ini dikenal sebagai *downsampling*, yaitu penyederhanaan data dengan tetap mempertahankan informasi utama yang relevan. Dengan mengurangi ukuran data, *pooling layer* dapat meningkatkan efisiensi komputasi serta membantu mengurangi risiko *overfitting* dalam proses pelatihan model. Selain itu, *pooling* juga membuat model lebih tahan terhadap variasi kecil pada posisi objek dalam citra [6].

Secara umum, terdapat dua jenis *pooling* yang paling sering digunakan, yaitu *max pooling* dan *average pooling*. *Max pooling* bekerja dengan mengambil nilai maksimum dari setiap area kecil pada *feature map*, sehingga mampu mempertahankan fitur yang paling dominan. Sementara itu, *average pooling* mengambil nilai rata-rata dari area tersebut, sehingga menghasilkan representasi fitur yang lebih halus. Kedua metode ini digunakan sesuai dengan kebutuhan model, di mana *max pooling* lebih sering digunakan karena lebih efektif dalam mempertahankan informasi penting dari citra [6]. Secara matematis, operasi *max pooling* dapat dinyatakan sebagai berikut:

$$P(i, j) = \max_{(m, n) \in R} X(i + m, j + n) \quad (2)$$

Sedangkan untuk *average pooling*:

$$P(i, j) = \frac{1}{|R|} \sum_{(m,n) \in R} X(i + m, j + n) \quad (3)$$

Keterangan:

X = *feature map* input

P = output hasil *pooling*

R = area/jendela *pooling* (misalnya 2×2)

(i, j) = posisi output

Rumus tersebut menunjukkan bahwa nilai output diperoleh dari proses pemilihan nilai maksimum atau rata-rata pada setiap area tertentu dari *feature map*. Selain itu, ukuran output dari *pooling layer* dapat dihitung menggunakan rumus berikut:

$$O = \frac{N - F}{S} + 1 \quad (4)$$

Keterangan:

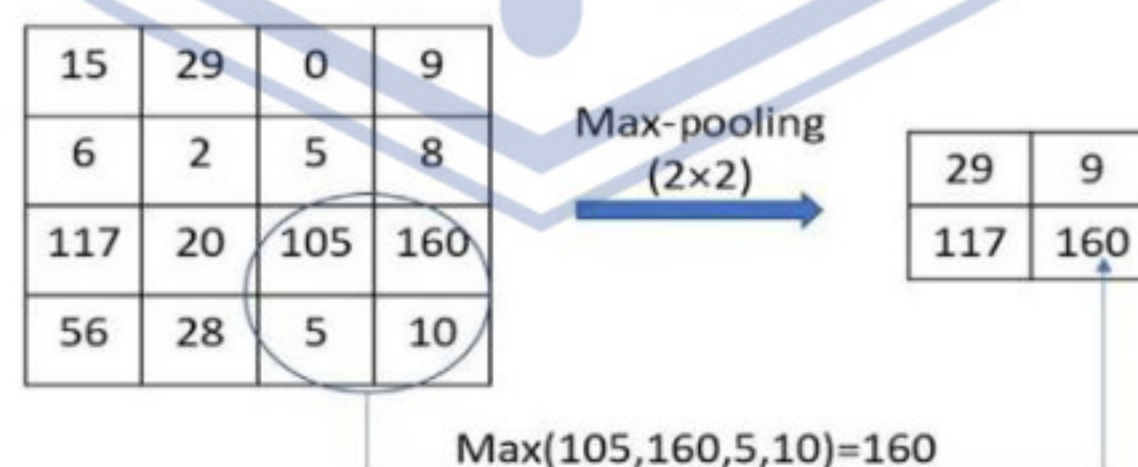
O = ukuran output

N = ukuran input

F = ukuran *pooling window*

S = *stride*

Rumus ini menunjukkan bahwa parameter seperti ukuran jendela dan *stride* akan mempengaruhi dimensi output yang dihasilkan. Dengan konfigurasi yang tepat, *pooling layer* mampu menghasilkan representasi fitur yang lebih ringkas namun tetap informatif. Berikut gambaran proses kerja dari *Pooling Layer* seperti di Gambar 2.13 dibawah ini.



Gambar 2. 13 *Pooling Layer* [5].

2.5.4 Fully Connected Layer

Fully Connected Layer (FCL) merupakan lapisan akhir dalam arsitektur *Convolutional Neural Network* (CNN) yang berfungsi untuk melakukan proses klasifikasi

berdasarkan fitur yang telah diekstraksi pada lapisan sebelumnya. Pada tahap ini, seluruh neuron pada lapisan sebelumnya dihubungkan secara penuh dengan neuron pada lapisan berikutnya, sehingga setiap fitur yang dihasilkan dari proses konvolusi dan *pooling* dapat digunakan secara maksimal dalam menentukan hasil prediksi. Sebelum masuk ke FCL, data *feature map* biasanya diubah terlebih dahulu menjadi vektor satu dimensi melalui proses *flattening* [6]. Secara matematis, proses pada *fully connected layer* dapat dinyatakan sebagai berikut:

$$W = [w_1, w_2, w_3, w_4, w_5, w_6], b_{fc} = [b_1, b_2, b_3, b_4, b_5, b_6] \quad (5)$$

Keterangan:

W =(Weight) Dalam operasi linear, W berperan sebagai koefisien atau pengali
 $w_1, w_2, \dots, w_6$ = Komponen Bobot Individu atau Sinapsis

b_{fc} = Bias di *Fully Connected*, yaitu nilai pergeseran fungsi dari titik pusat (0,0)

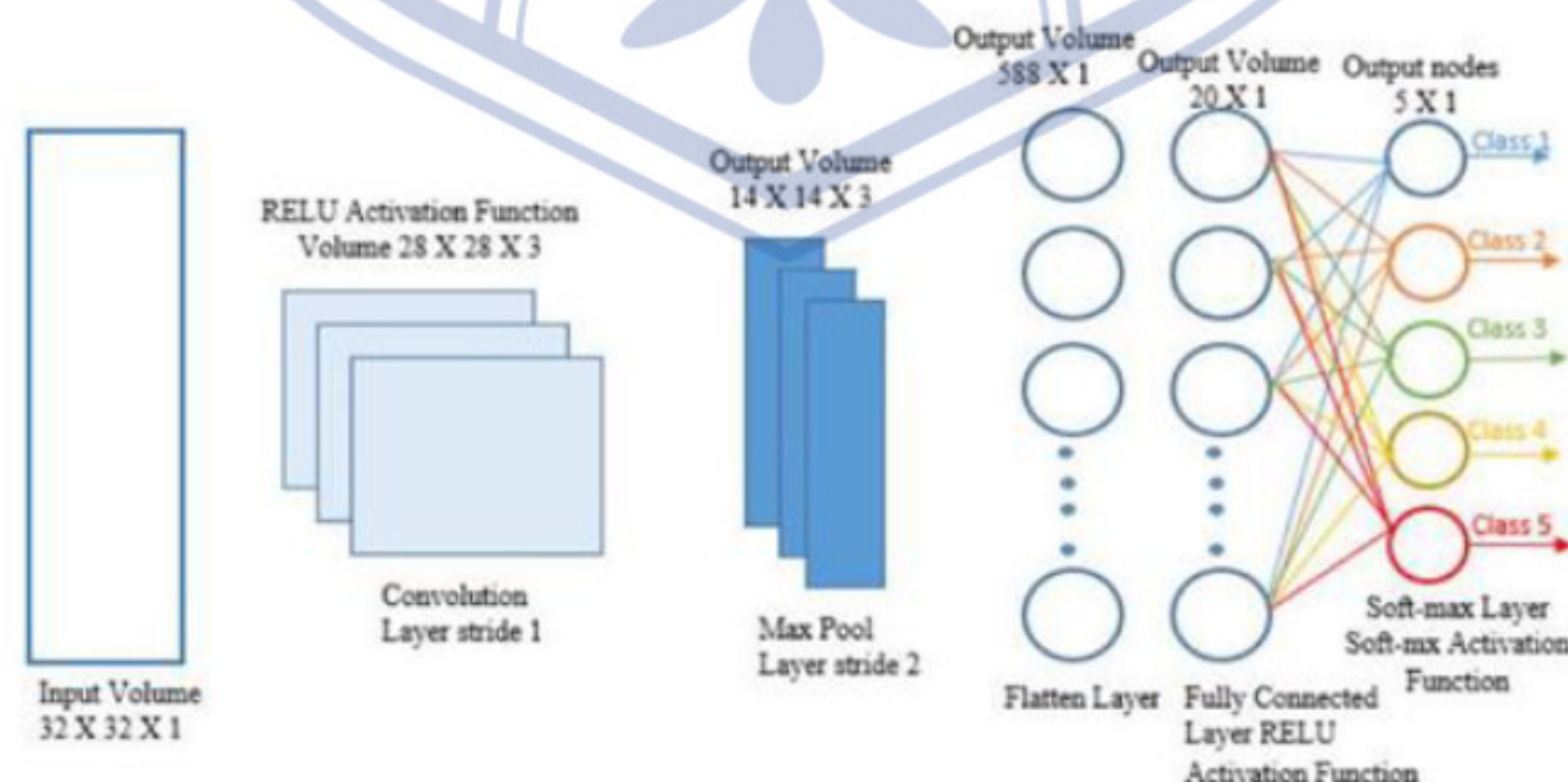
$b_1, b_2, \dots, b_6$ = Sebagai komponen penyusun di dalam vektornya masing-masing

Hasil perhitungan tersebut kemudian dilewatkan ke fungsi aktivasi untuk menghasilkan output akhir:

$$a_j = f(z_j) \quad (6)$$

di mana f merupakan fungsi aktivasi seperti ReLU, sigmoid, atau *softmax*. Pada kasus klasifikasi, fungsi *softmax* sering digunakan untuk menghasilkan probabilitas dari setiap kelas.

2.5.5 Fungsi Aktivasi



Gambar 2. 14 *Fully Connected Layer* [5].

Pada Gambar 2.14 di atas merupakan sebuah definisi dari Fungsi aktivasi yang merupakan komponen penting dalam *Convolutional Neural Network* (CNN) yang berfungsi untuk memberikan sifat non-linear pada model. Tanpa fungsi aktivasi, jaringan saraf hanya akan menghasilkan hubungan linear sehingga tidak mampu menangani pola kompleks pada data citra. Dengan adanya fungsi aktivasi, model dapat mempelajari hubungan non-linear yang lebih kompleks dan meningkatkan kemampuan dalam proses klasifikasi maupun deteksi pola. Beberapa fungsi aktivasi yang umum digunakan dalam CNN antara lain ReLU (*Rectified Linear Unit*), sigmoid, dan *softmax* [6].

1. ReLU (*Rectified Linear Unit*)

ReLU merupakan fungsi aktivasi yang paling umum digunakan dalam CNN karena sederhana dan efektif dalam mengatasi masalah *vanishing gradient*. Fungsi ini hanya mengubah nilai negatif menjadi nol, sementara nilai positif tetap dipertahankan.

$$A_{i,j} = \max(0, Z_{i,j}) \quad (7)$$

Grafik ReLU berbentuk garis lurus pada nilai positif dan nol pada nilai negatif, sehingga mempercepat proses pelatihan model.

2. Sigmoid

Fungsi sigmoid digunakan untuk mengubah nilai input menjadi rentang antara 0 dan 1, sehingga sering digunakan pada kasus klasifikasi biner.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Grafik sigmoid berbentuk kurva “S” yang halus, di mana nilai input kecil akan mendekati 0 dan nilai input besar akan mendekati 1. Meskipun mudah diinterpretasikan sebagai probabilitas, fungsi sigmoid memiliki kelemahan yaitu rentan terhadap masalah *vanishing gradient*, sehingga kurang efektif digunakan pada lapisan tersembunyi dalam jaringan yang dalam (*deep network*). Oleh karena itu, sigmoid umumnya digunakan pada lapisan output untuk kasus klasifikasi biner.

3. *Softmax*

Softmax merupakan fungsi aktivasi yang digunakan pada lapisan output untuk klasifikasi multi-kelas. Fungsi ini mengubah nilai input menjadi probabilitas dengan total jumlah sama dengan 1.

$$F\hat{y}_c = \frac{e^{f_c}}{\sum_{j=1}^n e^{x_j}} \quad (9)$$

Fungsi *softmax* memungkinkan model untuk menentukan probabilitas setiap kelas sehingga hasil klasifikasi dapat diinterpretasikan dengan lebih jelas.

Secara keseluruhan, pemilihan fungsi aktivasi sangat berpengaruh terhadap performa model CNN. ReLU sering digunakan pada lapisan tersembunyi karena efisien dan cepat, sigmoid digunakan untuk klasifikasi biner, sedangkan *softmax* digunakan untuk klasifikasi multi-kelas. Kombinasi yang tepat dari fungsi-fungsi ini akan meningkatkan kemampuan model dalam mengenali pola kompleks, termasuk dalam deteksi *deepfake*.

2.5.6 Fungsi *Loss* dan *Optimizer*

Fungsi *loss* (atau *loss function*) merupakan komponen penting dalam pelatihan *Convolutional Neural Network* (CNN) yang digunakan untuk mengukur seberapa besar kesalahan antara hasil prediksi model dengan nilai sebenarnya (*ground truth*). Nilai *loss* yang dihasilkan menjadi acuan bagi model untuk melakukan perbaikan selama proses pelatihan. Semakin kecil nilai *loss*, maka semakin baik performa model dalam melakukan prediksi. Fungsi ini bekerja dengan membandingkan output model dengan label sebenarnya, kemudian menghasilkan nilai numerik yang menunjukkan tingkat kesalahan [6]. Salah satu fungsi *loss* yang paling umum digunakan dalam klasifikasi adalah *cross-entropy loss*. Untuk klasifikasi biner, digunakan *binary cross-entropy* dengan rumus:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (10)$$

Sedangkan untuk klasifikasi multi-kelas, digunakan *categorical cross-entropy*:

$$L = -\sum_{i=1}^C y_i \log(\hat{y}_i) \quad (11)$$

Keterangan:

y_i = label sebenarnya

$\hat{y}_i$ = probabilitas prediksi model

N = jumlah data

C = jumlah kelas

Rumus tersebut menunjukkan bahwa semakin jauh nilai prediksi dari nilai sebenarnya, maka nilai *loss* akan semakin besar. Selain fungsi *loss*, proses pelatihan CNN juga membutuhkan *optimizer*, yaitu algoritma yang digunakan untuk memperbarui bobot (*weights*) dalam jaringan agar nilai *loss* dapat diminimalkan. Salah satu metode dasar yang sering digunakan adalah *Gradient Descent*, dengan rumus:

$$w = w - \alpha \frac{\partial L}{\partial w} \quad (12)$$

Keterangan:

w = bobot

α = *learning rate*

$\frac{\partial L}{\partial w}$ = gradien terhadap bobot

Selain itu, terdapat *optimizer* yang lebih canggih seperti **Adam (*Adaptive Moment Estimation*)** yang menggabungkan konsep *momentum* dan *adaptive learning rate*, sehingga proses pelatihan menjadi lebih cepat dan stabil.

2.6 Hyperparameter Tuning Grid Search

Grid Search merupakan salah satu metode *hyperparameter optimization* yang digunakan untuk memperoleh kombinasi *hyperparameter* terbaik pada model *machine learning* maupun *deep learning*. Metode ini bekerja dengan cara mendefinisikan sejumlah nilai kandidat untuk setiap *hyperparameter*, kemudian mencoba seluruh kombinasi yang mungkin secara sistematis (*exhaustive search*). Setiap kombinasi *hyperparameter* akan digunakan untuk melatih model dan dievaluasi menggunakan metrik tertentu, seperti *validation accuracy*, *validation loss*, *AUC*, atau metrik evaluasi lainnya. Kombinasi *hyperparameter* yang menghasilkan nilai evaluasi terbaik kemudian dipilih sebagai konfigurasi optimal model. Pendekatan ini bersifat deterministik karena seluruh kombinasi yang telah ditentukan akan dievaluasi tanpa ada yang terlewat [39].

Grid Search banyak digunakan karena sederhana, mudah diimplementasikan, dan mampu menghasilkan konfigurasi *hyperparameter* yang optimal ketika jumlah parameter yang diuji relatif sedikit. Pada model *Convolutional Neural Network* (CNN), metode ini umumnya digunakan untuk mengoptimalkan jumlah *filter*, ukuran *kernel*, jumlah neuron pada *fully connected layer*, *dropout rate*, *learning rate*, *batch size*, maupun jumlah *epoch*. Meskipun demikian, kelemahan utama *Grid Search* adalah tingginya kebutuhan komputasi karena seluruh kombinasi *hyperparameter* harus dievaluasi. Seiring bertambahnya jumlah

hyperparameter dan nilai kandidat yang diuji, jumlah kombinasi akan meningkat secara eksponensial sehingga waktu pelatihan menjadi lebih lama. Oleh karena itu, *Grid Search* lebih sesuai digunakan pada ruang pencarian (*search space*) yang tidak terlalu besar atau ketika peneliti telah menentukan rentang *hyperparameter* yang spesifik [40].

2.7 Evaluasi Kinerja Model

Evaluasi kinerja model merupakan tahapan krusial dalam siklus pengembangan kecerdasan buatan untuk memastikan bahwa sistem mampu memberikan hasil prediksi yang valid dan reliabel. Dalam berbagai riset informatika, *Confusion Matrix* digunakan sebagai alat ukur utama untuk mengevaluasi model klasifikasi dengan membandingkan nilai prediksi terhadap nilai aktual pada *dataset*. Matriks ini memberikan gambaran visual yang jelas mengenai keberhasilan model dalam mengenali objek, baik itu berupa wajah asli, konten *deepfake*, maupun denominasi uang kertas. Tanpa adanya evaluasi yang terstruktur, pengembang tidak dapat mengetahui secara pasti apakah algoritma yang dibangun sudah optimal atau masih memerlukan penyesuaian pada parameter tertentu [10]. Melalui matriks ini, performa lapisan *Fully Connected* dan fungsi aktivasi *Softmax* dalam memetakan fitur kompleks menjadi keputusan akhir dapat diukur secara kuantitatif.

a. Akurasi (*Accuracy*)

Parameter pertama yang sering menjadi rujukan adalah Akurasi (*Accuracy*), yang merepresentasikan proporsi prediksi benar dibandingkan dengan total keseluruhan data. Pada penelitian [7] pendeteksian nominal uang, akurasi tinggi sangat dipengaruhi oleh kualitas input yang telah melalui tahap prapemrosesan citra seperti konversi *grayscale* dan reduksi *noise*. Meskipun akurasi memberikan gambaran umum, nilai ini terkadang bisa menyesatkan jika *dataset* yang digunakan memiliki jumlah kelas yang tidak seimbang. Oleh karena itu, akurasi biasanya dipadukan dengan metrik lain untuk mendapatkan analisis performa yang lebih objektif dan mendalam.

b. Presisi (*Precision*)

Presisi (*Precision*) menjadi metrik penting berikutnya yang mengukur tingkat keakuratan model dalam memprediksi kelas positif. Dalam konteks deteksi *deepfake*, presisi tinggi memastikan bahwa video yang diidentifikasi sebagai manipulasi oleh sistem memang benar-benar merupakan konten palsu [41]. Hal ini sangat vital untuk menghindari kesalahan fatal berupa pelabelan negatif terhadap konten asli yang dapat merugikan reputasi individu. Dengan mengoptimalkan presisi melalui pemilihan arsitektur CNN yang tepat, sistem dapat meminimalisir angka *False Positive* atau kesalahan deteksi yang tidak semestinya.

c. Sensitivitas (*Recall*)

Di sisi lain, Sensitivitas atau *Recall* berfungsi untuk mengukur sejauh mana model mampu menemukan kembali seluruh data positif yang ada dalam *dataset*. Pada sistem deteksi marka jalan, *Recall* yang tinggi menjamin bahwa setiap garis marka dapat terdeteksi oleh sistem meskipun kondisi jalan sedang mengalami gangguan visual atau pencahayaan minim [7]. Ketidakmampuan model dalam mencapai nilai *Recall* yang optimal dapat menyebabkan informasi penting terlewatkan, yang dalam sistem navigasi bisa berakibat pada kegagalan deteksi objek vital. Oleh karena itu, keseimbangan antara Presisi dan *Recall* selalu menjadi fokus utama dalam pelatihan jaringan saraf.

d. *F1-Score*

Sebagai solusi untuk menyeimbangkan antara Presisi dan *Recall*, digunakanlah *F1-Score* yang merupakan rata-rata harmonik dari kedua metrik tersebut. *F1-Score* memberikan nilai representatif yang lebih stabil, terutama ketika menghadapi *dataset* dengan distribusi kelas yang sangat bervariasi. Dalam pengujian model klasifikasi wajah menggunakan metode CNN, *F1-Score* sering dijadikan indikator utama untuk menyatakan bahwa model memiliki performa yang seimbang dalam mengenali wajah asli maupun buatan. Penggunaan metrik ini memastikan bahwa performa model tidak condong hanya pada salah satu parameter saja [10].

e. AUC (*Area Under the Curve*)

AUC (*Area Under the Curve*) merupakan metrik yang digunakan untuk mengukur kemampuan model dalam membedakan 2 kelas pada berbagai nilai *threshold*. AUC diperoleh dari luas area di bawah kurva ROC (*Receiver Operating Characteristic*) yang menggambarkan hubungan *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). Pada penelitian deteksi *deepfake*, AUC digunakan untuk mengetahui kemampuan model dalam membedakan citra asli (*real*) dan citra hasil manipulasi (*deepfake*) secara keseluruhan. Nilai AUC yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan diskriminasi yang semakin baik, sedangkan nilai yang mendekati 0,5 menunjukkan bahwa kemampuan model tidak jauh lebih baik dibandingkan prediksi secara acak.

Implementasi nyata dari evaluasi ini terlihat pada pengujian sistem pendeteksian uang kertas yang menunjukkan bahwa integrasi antara teknik prapemrosesan citra dan arsitektur CNN mampu menghasilkan tingkat klasifikasi yang sangat presisi melalui dataset ribuan sampel [7]. Selain itu, evaluasi model juga berperan besar dalam menghadapi tantangan keamanan siber akibat penyalahgunaan teknologi *deepfake*. Dengan menggunakan *Confusion Matrix*, peneliti dapat memvalidasi efektivitas penggunaan *Vision Transformers*

yang dipadukan dengan CNN dalam mengenali anomali pada fitur mata dan hidung [41].

Sebagai kesimpulan, penggunaan *Confusion Matrix* yang mencakup Akurasi, Presisi, *Recall*, dan *F1-Score* merupakan standar wajib dalam pengembangan aplikasi kecerdasan buatan modern untuk memberikan diagnosa komprehensif terhadap kemampuan model dalam menangani data dunia nyata yang kompleks. Evaluasi kinerja yang transparan dan rinci pada akhirnya akan membangun kepercayaan publik terhadap penggunaan teknologi AI di berbagai lini kehidupan. Berikut adalah ringkasan hasil evaluasi model berdasarkan data referensi yang telah dibahas yang terdapat pada Tabel 2.1 dibawah ini.

Tabel 2. 1 Tabel Evaluasi Kinerja Model

Metrik Evaluasi	Rumus Perhitungan	Fungsi Utama
Akurasi	$\frac{TP + TN}{TP + TN + FP + FN}$	Mengukur kedekatan prediksi total dengan nilai aktual
Presisi	$\frac{TP}{TP + FP}$	Memastikan ketepatan deteksi konten manipulasi (<i>deepfake</i>)
<i>Recall</i>	$\frac{TP}{TP + FN}$	Menjamin tidak adanya konten hoaks visual yang lolos deteksi
<i>F1-Score</i>	$2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$	Menyeimbangkan performa pada <i>dataset</i> yang tidak seimbang
<i>AUC</i>	$AUC = \int_0^1 TPR(FPR)d(FPR)$	Mengukur kemampuan model dalam membedakan citra <i>real</i> dan <i>deepfake</i> pada berbagai nilai <i>threshold</i>

(Keterangan: TP = *True Positive*, TN = *True Negative*, FP = *False Positive*, FN = *False Negative*)

Berdasarkan Tabel 2.1 di atas, Akurasi merupakan metrik paling dasar yang memberikan gambaran umum tentang seberapa sering model berhasil memberikan prediksi yang benar terhadap *dataset* wajah asli dan buatan secara total [10]. Meskipun akurasi tinggi

menunjukkan kemampuan generalisasi yang baik, metrik ini dapat menjadi menipu jika digunakan pada *dataset* yang tidak seimbang (*imbalanced data*). Oleh karena itu, akurasi perlu dipadukan dengan metrik lain agar peneliti mendapatkan analisis performa yang lebih objektif dan mendalam dalam memastikan keandalan sistem deteksi.

Dalam berbagai penelitian berbasis *Convolutional Neural Network* (CNN), performa model umumnya dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, AUC, dan *loss*. Namun, hingga saat ini belum terdapat standar baku yang menetapkan batas numerik tertentu untuk menyatakan suatu model memiliki performa yang baik. Oleh karena itu, interpretasi hasil evaluasi biasanya dilakukan dengan membandingkan performa model terhadap penelitian terdahulu yang menggunakan dataset dan permasalahan yang serupa. Selain itu, penggunaan beberapa metrik secara bersamaan lebih disarankan dibandingkan hanya menggunakan *accuracy* karena setiap metrik memberikan informasi yang berbeda mengenai kemampuan model dalam melakukan klasifikasi.

Berdasarkan beberapa penelitian klasifikasi citra menggunakan CNN, model yang memperoleh nilai *accuracy*, *precision*, dan *recall* di atas 90% umumnya dilaporkan memiliki performa yang sangat baik (*high-performance model*). Sebagai contoh, penelitian pada klasifikasi citra yang membandingkan beberapa arsitektur CNN menunjukkan bahwa model dengan *accuracy* sebesar 93,25%, *precision* 99,05%, dan *recall* 87,25% sudah dianggap memberikan peningkatan performa yang signifikan dibandingkan model CNN dasar. Penelitian tersebut juga menunjukkan bahwa peningkatan *accuracy* biasanya diikuti oleh peningkatan *precision* dan *recall* sehingga model mampu melakukan klasifikasi secara lebih konsisten [42].

2.8 Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah sebuah bahasa standar yang digunakan untuk menspesifikasikan, memvisualisasikan, membangun, dan mendokumentasikan artefak dari sistem perangkat lunak berbasis objek (*object-oriented programming*). Dalam pengembangan sistem deteksi citra *deepfake*, UML berfungsi sebagai media cetak biru (*blueprint*) yang memodelkan aliran data dan fungsionalitas sistem secara terstruktur sebelum diimplementasikan ke dalam kode program. Pemodelan ini memungkinkan arsitektur perangkat lunak mulai dari proses pengunggahan citra dari media sosial, tahap pra-proses dengan MTCNN, hingga klasifikasi menggunakan *Convolutional Neural Network* (CNN) dapat dipetakan dengan jelas, sehingga memudahkan analisis kebutuhan sistem dan interaksi antar komponen di dalamnya [43].

Secara umum, diagram UML dibagi menjadi dua kategori utama yang merepresentasikan sudut pandang sistem secara komprehensif, yaitu diagram struktural (*structural diagram*) dan diagram perilaku (*behavioral diagram*). Dalam penelitian ini, diagram perilaku seperti *Use Case Diagram* digunakan untuk menggambarkan interaksi antara aktor (pengguna atau sistem media sosial) dengan fungsionalitas utama aplikasi deteksi hoaks visual, sementara *Activity Diagram* dimanfaatkan untuk memodelkan urutan aktivitas atau alur kerja (*workflow*) dari algoritma CNN dalam memproses citra dari awal hingga menghasilkan status validasi. Melalui kombinasi diagram-diagram tersebut, kompleksitas alur komputasi kecerdasan buatan dapat disederhanakan ke dalam bentuk visual yang baku, logis, dan mudah dipahami dalam kerangka rekayasa perangkat lunak [43].

2.9 PIECES (*Performance, Information, Economics, Control, Efficiency, and Service*)

Metode analisis PIECES (*Performance, Information, Economics, Control, Efficiency, and Service*) merupakan sebuah kerangka kerja yang digunakan untuk mengevaluasi dan menganalisis kelemahan serta kebutuhan dari sebuah sistem berjalan maupun sistem yang akan dikembangkan [44]. Dalam penelitian deteksi citra *deepfake* berbasis *Convolutional Neural Network* (CNN) ini, analisis PIECES berfungsi sebagai landasan teoretis untuk memetakan urgensi pengembangan sistem dalam mendeteksi hoaks visual di media sosial. Melalui metode ini, peneliti dapat mengidentifikasi masalah dari sistem klasifikasi konvensional atau deteksi manual yang ada saat ini, sehingga dapat dirumuskan solusi teknologi yang lebih responsif dan akurat dalam membendung penyebaran konten manipulatif.

Kerangka analisis ini bekerja secara komprehensif dengan mengevaluasi enam dimensi utama sistem. Aspek *Performance* (Kinerja) menilai kecepatan dan ketepatan sistem dalam memproses serta mengklasifikasikan citra *deepfake*. *Information* (Informasi) berfokus pada kualitas dan validitas output informasi yang dihasilkan untuk pengguna. *Economics* (Ekonomi) meninjau efisiensi biaya dan sumber daya yang dibutuhkan dalam implementasi model CNN. *Control* (Pengendalian) mengevaluasi sistem keamanan dalam menyaring data dan mencegah celah kesalahan deteksi (*false alarms*). Sementara itu, aspek *Efficiency* (Efisiensi) mengukur optimalisasi penggunaan memori serta komputasi selama proses *training* dan *testing*, dan *Service* (Pelayanan) menilai tingkat kemudahan serta keandalan layanan deteksi tersebut ketika digunakan oleh masyarakat atau diintegrasikan ke platform media sosial untuk pencegahan hoaks visual [44].

2.10 Stratified Split

Teknik *stratified splitting* adalah metode pembagian *dataset* yang mempertahankan distribusi label secara proporsional antara subset pelatihan dan pengujian. Dalam dokumentasi resmi *Scikit-learn* (`train_test_split`), disarankan untuk menggunakan parameter `stratify=y` agar distribusi kelas tetap konsisten setelah pembagian data. Hal ini sangat krusial untuk klasifikasi, terutama ketika label tidak terdistribusi merata. Dokumentasi tersebut menyatakan bahwa tanpa stratifikasi, pembagian acak dapat menghasilkan subset pelatihan atau pengujian yang tidak mencerminkan proporsi kelas asli, yang pada akhirnya menyebabkan model bias terhadap kelas mayoritas. Oleh karena itu, *stratified sampling* menjadi praktik baku untuk menjaga performa model yang lebih representatif dan adil terhadap seluruh kelas, khususnya dalam *dataset* yang *imbalanced*. Berikut ini merupakan rumus dari *stratified split*:

$$n_c^{subset} \approx \left\lfloor \frac{n_c}{N} \times N_{subset} \right\rfloor \quad (13)$$

Dengan:

n_c = jumlah data untuk kelas c

N = total data seluruh kelas

N_{subset} = jumlah total data yang dialokasikan ke subset tertentu (*train*, *val*, & *test*)

N_c^{subset} = jumlah data kelas c yang akan dimasukkan ke subset tersebut

2.11 Penelitian Terdahulu

Berikut adalah penyusunan Penelitian Terdahulu yang mencakup fokus penelitian, metode yang digunakan, serta hasil atau kontribusi utamanya dari beberapa sumber referensi tersebut yang terdapat pada Tabel 2.1 dibawah ini..

Tabel 2. 2 Penelitian Terdahulu

No.	Peneliti & Tahun	Judul Penelitian	Metode / Arsitektur	Hasil & Kontribusi Utama
1.	RD. Kusumanto, dkk (2011)	<i>Pengolahan Citra Digital untuk Mendeteksi Obyek Menggunakan Pengolahan Warna Model Normalisasi RGB [26]</i>	Normalisasi RGB	Memberikan dasar matematis citra sebagai matriks $f(x, y)$ dan teknik segmentasi warna untuk deteksi obyek awal [26].

2.	Muhamad Malik Ibrahim, dkk (2023)	<i>Pendeteksian Nominal Uang pada Gambar Menggunakan Convolutional Neural Network [28]</i>	CNN & Pra-pemrosesan Citra	Berhasil mengklasifikasi 8 denominasi uang kertas menggunakan 1076 sampel data dengan akurasi tinggi melalui integrasi prapemrosesan [28]
3.	Damar Wicaksono, dkk (2024)	<i>Analisis Perbandingan Metode Pra Pemrosesan Citra untuk Deteksi Tepi Canny pada Berbagai Kondisi Jalan [29]</i>	Canny Edge Detection & Adaptive Thresholding	Mengoptimalkan deteksi marka jalan melalui filter noise dan thresholding adaptif untuk meningkatkan keamanan navigasi [29].
4.	Dede Latipah, dkk (2025)	<i>Analisis Implementasi Artificial Intelligence dalam Optimalisasi Proses Bisnis [31]</i>	Analisis Deskriptif & Statistik AI	Menunjukkan bahwa AI meningkatkan percepatan kerja sebesar 78% dan efisiensi biaya operasional sebesar 65% dalam dunia bisnis [31].
5.	Zulfa Putri, dkk (2025)	<i>Analisis Dampak Penyalahgunaan Teknologi Deepfake Terhadap Keamanan Digital Masyarakat di Indonesia [32]</i>	Studi Kasus & Tinjauan Yuridis	Menganalisis penyalahgunaan AI untuk penipuan finansial serta peran UU ITE & UU PDP dalam melindungi data pribadi masyarakat [32]

6.	Prastita, et al. (2025)	<i>Analisis Perbandingan Metode CNN untuk Deteksi Warna pada Objek [39]</i>	CNN (VGG16, Xception)	Membuktikan efektivitas metrik <i>Confusion Matrix</i> dalam membandingkan berbagai arsitektur CNN [39].
7.	Sakib, et al. (2018)	<i>An Overview of CNN: Its Architecture and Applications [40]</i>	Deep Learning CNN	Memberikan tinjauan komprehensif mengenai penerapan CNN dalam tugas pengenalan citra secara adaptif [40].
8.	Yamashita, et al. (2018)	<i>Convolutional neural networks: an overview and application [42]</i>	CNN & Backpropagation	Menjelaskan proses pembelajaran hierarki spasial melalui lapisan-lapisan saraf untuk ekstraksi fitur otomatis [42].
9.	Jesselyn Mu, dkk (2024)	<i>Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network [43]</i>	CNN & Confusion Matrix	Membangun model konseptual untuk membedakan wajah asli dan buatan dengan validasi kinerja menggunakan metrik matriks evaluasi [43].

10.	Ahmed Hatem Soudy, dkk (2024)	<i>Deepfake detection using convolutional vision transformers and convolutional neural networks</i> [44]	CNN & Vision Transformers (ViT)	Mengembangkan sistem deteksi <i>deepfake</i> tiga komponen (prapemrosesan, deteksi, prediksi) dengan fokus pada fitur mata dan hidung [44].
-----	-------------------------------	--	---------------------------------	---

Berdasarkan tabel penelitian terdahulu di atas, terlihat adanya benang merah yang jelas antara pengolahan citra dasar dengan perkembangan AI modern, di mana teknik prapemrosesan menjadi kunci utama bagi algoritma CNN dalam mengekstraksi fitur secara akurat untuk mendeteksi manipulasi *deepfake*. Tinjauan terhadap berbagai studi tersebut mengungkapkan beberapa poin fundamental sebagai berikut:

a) Evolusi Metode

Penelitian dalam bidang visi komputer telah mengalami transformasi yang signifikan, berkembang dari metode konvensional seperti segmentasi warna sederhana dan deteksi tepi mekanistik menuju penggunaan *Deep Learning* yang jauh lebih kompleks seperti arsitektur CNN dan *Transformers* [38]. Evolusi ini menunjukkan bahwa pemahaman model terhadap objek kini tidak lagi terbatas pada atribut fisik sederhana, melainkan sudah mampu mengenali fitur-fitur semantik yang mendalam pada citra digital [26].

b) Pentingnya Prapemrosesan

Seluruh penelitian teknis yang dirujuk secara konsisten menekankan bahwa kualitas akhir dari deteksi baik itu pada objek uang kertas, marka jalan, maupun fitur wajah sangat bergantung pada tahap pembersihan *noise* dan normalisasi citra di awal proses [26]. Tahap prapemrosesan yang optimal memastikan bahwa input yang diterima oleh lapisan konvolusi telah terstandarisasi, sehingga meminimalkan beban komputasi dan meningkatkan ketajaman ekstraksi fitur.

c) Evaluasi Objektif

Penelitian terbaru secara konsisten mengadopsi penggunaan *Confusion Matrix* sebagai instrumen validasi utama untuk memastikan hasil deteksi bukan sekadar kebetulan, melainkan memiliki akurasi, presisi, dan *recall* yang dapat dipertanggungjawabkan secara ilmiah. Hal ini sejalan dengan Evaluasi Kinerja Model, di mana metrik evaluasi menjadi standar wajib untuk menguji keandalan algoritma sebelum diimplementasikan pada data

dunia nyata yang kompleks.

d) Dampak Sosial & Bisnis

Selain aspek teknis, literatur juga menyoroti dimensi yang lebih luas mengenai bagaimana AI memberikan keuntungan ekonomi melalui otomatisasi, sekaligus menghadirkan tantangan etika dan hukum yang serius terkait manipulasi digital. Integrasi antara kekuatan teknologi CNN dalam mendeteksi *deepfake* dengan regulasi keamanan digital menjadi solusi komprehensif dalam melindungi masyarakat dari ancaman hoaks visual di media sosial [4].

