

BAB II

TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Jesselyn Mu el al. [6] berfokus pada pengembangan model konseptual untuk mengidentifikasi perbedaan antara citra wajah asli dan citra wajah hasil rekayasa *deepfake*. Penelitian ini dilatarbelakangi oleh pesatnya perkembangan *Artificial Intelligence* yang tidak hanya memberikan dampak positif, tetapi juga potensi penyalahgunaan, salah satunya melalui teknologi *deepfake* yang dapat merugikan individu maupun kelompok. Dalam penelitian ini, digunakan metode CNN dengan rancangan arsitektur yang terdiri atas enam *convolutional layer*, tiga *max pooling layer*, serta penambahan *batch normalization layer* yang dirancang menyerupai arsitektur *VGG*. CNN dipilih karena kemampuannya yang baik dalam mengekstraksi dan mengklasifikasikan fitur visual dari citra. Sumber dataset pada penelitian ini berasal dari *Kaggle* yang merupakan dataset publik. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, dan menunjukkan hasil yang cukup tinggi, yaitu sebesar 91% pada seluruh metrik evaluasi tersebut. Hal ini menunjukkan bahwa arsitektur yang dikembangkan memiliki potensi yang baik dalam membedakan citra wajah asli dan wajah hasil *deepfake*.

Penelitian yang dilakukan oleh M. Azka et al. [7] bertujuan untuk membandingkan performa pendeteksian *deepfake* menggunakan tiga arsitektur CNN, yaitu ResNet-50, VGG-19, dan Xception, yang dikombinasikan dengan *capsule networks*. Arsitektur *capsule networks* yang digunakan terdiri dari dua *convolutional layers* pada kapsul primer dan dua kapsul keluaran yang merepresentasikan label *real* dan *deepfake*. Dataset yang digunakan dalam penelitian berasal dari *Kaggle* yang terdiri dari 6.000 citra wajah asli dan 4.000 citra wajah hasil *deepfake*, dilatih selama 100 *epochs* dengan *batch size* 10. Evaluasi dilakukan dengan menguji 20 gambar acak untuk mengukur kemampuan klasifikasi masing-masing model. Hasil penelitian menunjukkan bahwa ketiga arsitektur memiliki performa deteksi yang bervariasi, dengan ResNet-50 mencapai akurasi sebesar 64,133%, Xception sebesar 64,067%, dan VGG-19 sebesar 61,067%. Hal

ini menunjukkan bahwa meskipun model CNN memiliki potensi dalam mendeteksi *deepfake*, tingkat akurasi masih dapat ditingkatkan, terutama melalui pemilihan arsitektur yang lebih optimal atau pendekatan yang lebih adaptif terhadap kompleksitas fitur visual pada citra *deepfake*.

Penelitian yang dilakukan oleh Ananda Adhicitta [9] bertujuan untuk mengembangkan dan menerapkan *deepfake detector* guna mengenali informasi palsu berbentuk citra wajah yang tersebar di media sosial. Penelitian ini memanfaatkan algoritma *deep learning* berbasis *Convolutional Neural Network*, secara khusus menggunakan arsitektur ResNet-50, untuk mendeteksi gambar hoaks yang dihasilkan oleh model *Generative Adversarial Networks* (GAN), terutama ProGAN. Dataset pelatihan pada penelitian ini tidak disebutkan secara detail, namun dataset yang digunakan merupakan dataset publik. Dalam eksperimennya, model ResNet-50 dilatih untuk mengenali pola-pola halus yang tidak kasat mata pada citra yang dimanipulasi secara digital. Metrik evaluasi yang digunakan meliputi akurasi, presisi, dan *recall*, dengan hasil yang diperoleh yaitu akurasi sebesar 85%, presisi sebesar 100%, dan *recall* sebesar 65%. Meskipun performa detektor cukup tinggi terhadap gambar yang dibuat menggunakan ProGAN, performa model menurun ketika digunakan untuk mendeteksi gambar dari model GAN lain seperti StyleGAN dan BigGAN. Hal ini mengindikasikan pentingnya pengembangan arsitektur atau strategi pelatihan yang lebih generalis agar model mampu mengenali variasi teknik manipulasi *deepfake* yang beragam.

Penelitian oleh Muhammad Indra Abidi et al. [13] bertujuan untuk mendeteksi *deepfake* pada video menggunakan kombinasi arsitektur *ResNext Convolutional Neural Network* dan *Long Short-Term Memory* (LSTM). Penelitian ini memanfaatkan *computer vision* untuk menangani permasalahan pemalsuan wajah dalam video, yang dapat berdampak serius pada privasi individu. Dataset yang digunakan terdiri dari 200 video yang telah melalui proses *preprocessing* dengan pengubahan ukuran citra menjadi 112×112 piksel dan pengelompokan berdasarkan jumlah *frame* (10, 20, 40, dan 60). Arsitektur *ResNext CNN* digunakan untuk mengekstraksi fitur visual dari citra, sementara *LSTM* digunakan untuk

mengenali pola temporal dari data urutan video. Evaluasi kinerja sistem dilakukan menggunakan *confusion matrix*, dan hasil penelitian menunjukkan bahwa akurasi tertinggi sebesar 90% diperoleh pada video dengan 60 *frames*, sedangkan akurasi terendah sebesar 52% terdapat pada video dengan 10 *frames*. Hasil ini membuktikan bahwa kombinasi *ResNext CNN* dan *LSTM* efektif dalam mendeteksi video *deepfake* meskipun pada citra beresolusi kecil.

Penelitian oleh Anuj Badale et al. [14] dengan tujuan untuk mendeteksi video *deepfake* menggunakan pendekatan berbasis *neural networks*. Dalam penelitian ini, klasifikasi biner dilakukan menggunakan kombinasi lapisan *dense* dan *convolutional neural network* untuk membedakan antara video asli dan video hasil rekayasa. Meskipun detail dataset tidak disebutkan secara eksplisit, proses pelatihan model dilakukan dengan menggunakan fungsi kehilangan *categorical cross entropy*, *binary cross entropy*, dan *mean square error* yang masing-masing dikombinasikan dengan dua jenis *optimizer*, yaitu *Adam* dan *stochastic gradient descent (SGD)*. Dataset yang digunakan dalam penelitian ini berasal dari *FaceForensics*, yaitu sebuah himpunan data berskala besar yang dikembangkan oleh Rössler et al. (2018), dengan jumlah 900 video *deepfake*. Evaluasi kinerja sistem dilakukan dengan mengukur tingkat akurasi, di mana hasil terbaik diperoleh dengan *Adam optimizer* pada *categorical cross entropy* sebesar 91%. Penelitian ini menunjukkan bahwa kombinasi lapisan *dense* dan CNN cukup efektif dalam melakukan klasifikasi biner untuk mendeteksi video *deepfake*.

Penelitian yang dilakukan oleh Dewi et al. [16] bertujuan untuk membangun sistem klasifikasi citra wajah palsu guna membedakan antara gambar asli (*real*) dan gambar hasil manipulasi digital (*fake*) yang sering disalahgunakan di *media sosial*. Penelitian ini memanfaatkan arsitektur *FaceNet* untuk ekstraksi fitur wajah dan menggabungkannya dengan metode *Error Level Analysis* guna mendeteksi tingkat rekayasa pada citra. Data yang digunakan dalam penelitian ini merupakan gabungan data primer hasil pengambilan dan penyuntingan citra mandiri, serta data sekunder dari situs *Kaggle* yang terdiri dari total 8.912 gambar, masing-masing terbagi seimbang dalam kelas *real* dan *fake*. Proses klasifikasi dilakukan menggunakan

algoritma *Support Vector Machines* (SVM) dengan penyetelan *hyperparameter* berupa *epoch* sebanyak 10, *optimizer Adam*, *batch size* sebesar 64, dan *learning rate* sebesar 0.0001. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil penelitian menunjukkan bahwa model mencapai akurasi sebesar 95,78%, dengan konfigurasi *hyperparameter* terbaik pada SVM menggunakan *kernel Linear* dan nilai *C* sebesar 0.1 yang menghasilkan *precision* sebesar 53%, *recall* 54%, dan *f1-score* sebesar 54%. Penelitian ini menunjukkan potensi penggabungan teknik deteksi manipulasi dan klasifikasi berbasis *machine learning* dalam menghadapi tantangan keamanan visual digital.

Kurniawan et al. [17] melakukan penelitian untuk mendeteksi citra wajah asli (*real*) dan hasil manipulasi deepfake pada data gambar dan video menggunakan algoritma *You Only Look Once* versi 11 (*YOLO11*). Penelitian ini bertujuan mengembangkan model deteksi yang efektif untuk membedakan wajah asli dan deepfake guna mengurangi risiko penyebaran informasi palsu dan pencurian identitas digital. Dataset publik yang digunakan meliputi *EC2-DeepFake* dan *Deepfake Dataset* yang bersumber dari *Roboflow*, sementara evaluasi performa model dilakukan dengan metrik *mean Average Precision* pada threshold 50% (*mAP50*) dan rentang 50-95% (*mAP50-95*). Hasil evaluasi menunjukkan model mencapai akurasi sedang dengan nilai *mAP50* sebesar 68,2% dan *mAP50-95* sebesar 57,8%. Model menunjukkan *recall* tinggi sebesar 84,9%, yang menandakan kemampuan menangkap sebagian besar objek, namun *precision* yang lebih rendah yaitu 58,1% menunjukkan tantangan dalam menghindari prediksi palsu. Penelitian juga menemukan bahwa model mengalami kesulitan saat mengolah gambar non-fotografis dan dataset yang tidak seimbang. Temuan ini menekankan pentingnya penggunaan dataset yang lebih besar dan beragam serta peningkatan teknik pelatihan guna meningkatkan kinerja model. Meskipun terdapat keterbatasan, studi ini mengonfirmasi potensi algoritma *YOLO11* sebagai fondasi pengembangan sistem deteksi deepfake yang dapat meningkatkan keaslian media digital secara signifikan.

Keseluruhan penelitian menunjukkan bahwa *Convolutional Neural Network* (CNN) memiliki peran sentral dalam pengenalan wajah dan deteksi *deepfake*, dengan hasil yang sangat menjanjikan. Berbagai penelitian menunjukkan bahwa CNN mampu memberikan akurasi yang tinggi dalam mengidentifikasi dan mengklasifikasikan ekspresi wajah, bahkan dalam kondisi data yang kompleks dan beragam. Selain itu, penggunaan kombinasi CNN dengan teknik lain seperti *Long Short-Term Memory* (LSTM) dan algoritma pendukung lainnya juga meningkatkan kemampuan deteksi, memberikan hasil yang konsisten dan dapat diandalkan.

Keberhasilan CNN dalam banyak aplikasi ini menunjukkan bahwa metode ini adalah pilihan utama dalam pengembangan sistem pengenalan wajah dan deteksi *deepfake*. Dengan kemampuan untuk mengekstraksi fitur-fitur penting dari gambar dan video secara otomatis, CNN tidak hanya meningkatkan akurasi, tetapi juga efisiensi dalam proses deteksi. Oleh karena itu, penerapan CNN dalam penelitian dan aplikasi praktis di bidang *image classification* adalah langkah yang sangat baik dan dapat diharapkan untuk terus berkembang di masa depan. Berikut adalah ringkasan perbandingan setiap penelitian terdahulu dengan penelitian yang dilakukan oleh peneliti sekarang pada **Tabel 2.1**.

Tabel 2. 1 Tabel Perbandingan Penelitian Sebelumnya

Ref	Tahun	Topik Penelitian	Arsitektur CNN Digunakan	Sumber Dataset	Metrik Evaluasi	Hasil Penelitian
[6]	2024	Deteksi wajah asli vs deepfake	6 <i>conv</i> layers, 3 <i>max pooling</i> , <i>batch norm</i> (mirip VGG)	Kaggle, oleh <u>Manjil Karki</u> (100.000+ dataset)	<i>Accuracy</i> , <i>Precision</i> , <i>Recall</i> , <i>F1-score</i>	91% pada semua metrik dan menghasilkan model
[7]	2023	Perbandingan CNN + CapsuleNet untuk deteksi deepfake	ResNet-50, VGG-19, Xception + CapsuleNet	Kaggle, sumber detail tidak disebutkan (6000 <i>real class</i> dan 400 <i>fake class</i>)	<i>Accuracy</i>	ResNet-50: 64,13%; Xception: 64,07%; VGG-19: 61,07%, menghasilkan model dan API
[9]	2023	Deteksi deepfake dari ProGAN	ResNet-50	Dataset publik, sumber tidak disebutkan dan jumlah dataset tidak disebutkan.	<i>Accuracy</i> , <i>Precision</i> , <i>Recall</i>	Accuracy: 85%; Precision: 100%; Recall: 65%, menghasilkan model dan <i>web app</i>
[13]	2023	Deteksi video deepfake menggunakan CNN + LSTM	ResNext CNN + LSTM	Sumber Dataset tidak disebutkan dan jumlah dataset tidak disebutkan.	<i>Accuracy</i> (<i>Confusion matrix</i>)	90% (60 frames), 52% (10 frames) dan menghasilkan model
[14]	2023	Deteksi video deepfake	Dense + CNN	<i>Faceforensics</i> , oleh Rössler et al. (900 video <i>deepfake</i>)	<i>Accuracy</i>	Accuracy tertinggi: 91% (Adam + <i>categorical</i>

						<i>cross entropy</i>) dan menghasilkan model
[16]	2023	Membangun sistem klasifikasi citra wajah <i>deepfake</i>	FaceNet + Support Vector Machines (SVM)	Kaggle dan punyuntingan mandiri, sumber detail tidak disebutkan (10.000++ dataset)	<i>Accuracy, Precision, Recall, F1-score</i>	Accuracy = 95,78%, Precision = 53%, Recall = 54%, F1-score = 54% dan menghasilkan model
[17]	2023	Mendeteksi citra wajah asli (<i>real</i>) dan hasil manipulasi <i>deepfake</i>	YOLO11	Roboflow, jumlah dataset tidak disebutkan.	<i>mAP50, mAP50-95, Precision, Recall</i>	mAP50 = 68,2%, mAP50-95 = 57,8%, Precision = 58,1%, Recall = 84,9% dan model
-	2025	Pendeteksi <i>deefake</i> pada gambar citra wajah	MobileNetV2, ResNet50, AlexNet	OpenForensics [26], 140.002 untuk pelatihan, 10.905 untuk validasi, dan 39.428.	<i>Accuracy, Precision, Recall, F1 - Score</i>	-