Jurnal Riset Aplikasi Ilmu Data dan Sistem 1 (1), 107–116 DOI: 0000
Juli 2026

Deteksi Tindakan Kecurangan Pada Ujian Daring Menggunakan ResNet50 dan *Bidirectional Long Short-Term Memory* (BiLSTM)

Valentino Prasetyo Putra[1], Elly Matul Imah[2,*]

[1] *Program Studi Sarjana Sains Data, Universitas Negeri Surabaya, Indonesia*
[2] *Program Studi Sarjana Kecerdasan Artifisial, Universitas Negeri Surabaya, Indonesia*

Abstrak

Pelaksanaan ujian daring menghadapi berbagai tantangan, salah satunya adalah potensi terjadinya kecurangan akibat keterbatasan pengawasan secara langsung. Penelitian ini mengusulkan pendekatan berbasis *deep learning* untuk mendeteksi dan mengklasifikasikan perilaku kecurangan selama ujian daring ke dalam enam kategori menggunakan data video *wearcam* dari *Online Exam Proctoring* (OEP) Dataset. Fitur visual diekstraksi menggunakan ResNet50 yang memanfaatkan mekanisme *residual learning* untuk menghasilkan representasi fitur dari setiap *frame* video. Fitur yang diekstraksi selanjutnya dimodelkan secara temporal menggunakan *Bidirectional Long Short-Term Memory* (BiLSTM) untuk menangkap dependensi waktu dari dua arah sekaligus. Berdasarkan hasil pengujian, kombinasi ResNet50 dan BiLSTM menghasilkan kinerja terbaik pada klasifikasi enam kelas dengan nilai *accuracy* sebesar 0.93, rata-rata *F1-score* sebesar 0.85, rata-rata *specificity* sebesar 0.98, dan rata-rata *sensitivity* sebesar 0.80. Temuan penelitian ini menunjukkan bahwa kombinasi ResNet50 dan BiLSTM mampu memberikan performa deteksi kecurangan yang efektif pada ujian daring dengan enam kategori perilaku.

Kata Kunci *deteksi kecurangan; deep learning; BiLSTM; ResNet50; ujian daring*

Abstract

Online examinations face various challenges, one of which is the increased risk of cheating due to the lack of direct supervision. This study proposes a deep learning approach for detecting and classifying cheating behavior during online examinations into six categories using wearcam video data from the Online Exam Proctoring (OEP) Dataset. Visual features were extracted using ResNet50 to generate feature representations from each video frame. The extracted features were then modeled temporally using Bidirectional Long Short-Term Memory (BiLSTM) to capture time dependencies from both directions. Experimental results showed that the ResNet50 and BiLSTM combination achieved the best performance in the six-class classification scenario, yielding an accuracy of 0.93, an average F1-score of 0.85, an average specificity of 0.98, and an average sensitivity of 0.80. These findings demonstrate that the combination of ResNet50 and BiLSTM provides effective cheating detection performance for online examination scenarios with six behavioral categories.

Keywords *cheating detection; deep learning; BiLSTM; ResNet50; online examination*

*Corresponding author. Email: ellymatul@unesa.ac.id.

Received 26 Juni 2026; Revised 11 Juli 2026; Accepted 12 Juli 2026

1 Pendahuluan

Pendidikan berperan penting dalam meningkatkan kualitas sumber daya manusia dan kemampuan adaptasi terhadap perkembangan teknologi. Perkembangan teknologi mendorong kemajuan metode dan media pembelajaran (Lindín dkk., 2023), baik yang dilakukan secara daring maupun luring, yang kini diadopsi oleh institusi formal maupun layanan pendidikan non-formal seperti kursus daring. Kemudahan akses, fleksibilitas, serta keberagaman materi menjadikan pembelajaran daring semakin diminati (Child dkk., 2023; Turan dkk., 2022). Dalam proses pembelajaran, evaluasi melalui ujian masih banyak digunakan sebagai sarana utama untuk mengukur pemahaman dan kompetensi peserta didik (Ababneh dkk., 2022). Integritas ujian daring memiliki tantangan tersendiri karena pengawasan tidak dapat dilakukan secara langsung sehingga ruang terjadinya kecurangan jauh lebih besar dibandingkan ujian luring (Fendler dkk., 2024). Newton dan Essex (2024) menunjukkan bahwa praktik kecurangan terjadi dalam skala signifikan, dengan lebih dari 44,7% mahasiswa mengaku pernah melakukannya dan angkanya meningkat selama periode pandemi. Kondisi ini menegaskan perlunya sistem pemantauan yang mampu mendeteksi perilaku kecurangan secara otomatis dalam ujian daring (Sakhipov dkk., 2025).

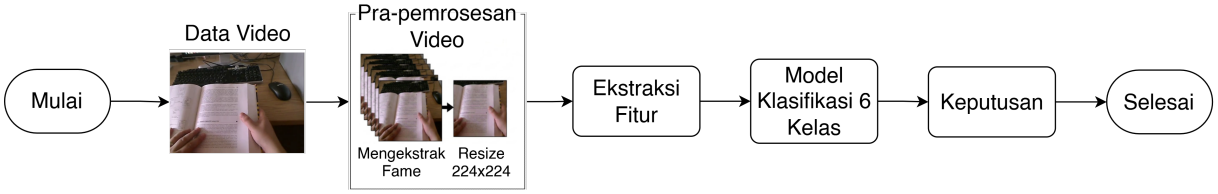
Berbagai penelitian telah mengusulkan pendekatan untuk deteksi kecurangan berbasis *deep learning*. Kaddoura dan Gumaei (2022) mengembangkan sistem yang memanfaatkan data visual untuk mendeteksi perilaku mencurigakan selama ujian daring. Pendekatan serupa juga dikembangkan oleh Lamba dan Sharma (2024), yang mengintegrasikan deteksi wajah, serta analisis arah pandangan dan pose kepala menggunakan arsitektur CNN dan BiLSTM untuk menghasilkan prediksi kecurangan yang lebih komprehensif. Essahraui dkk. (2025) berfokus pada analisis citra atau video menggunakan model berbasis *Convolutional Neural Network* (CNN) seperti ResNet50, MobileNetV2, dan EfficientNet serta metode deteksi objek. Pendekatan berbasis *machine learning* konvensional seperti *Support Vector Machine* (SVM), *Decision Tree*, dan XGBoost juga telah dieksplorasi untuk memprediksi perilaku curang pada ujian digital (Erdem dan Karabatak, 2025).

Penelitian yang dilakukan oleh Smirani dan Boulahia (2022) mengembangkan sistem deteksi kecurangan berbasis CNN yang menganalisis aktivitas peserta ujian melalui tangkapan layar komputer, pola navigasi, dan alamat IP tanpa menggunakan kamera. Meskipun pendekatan-pendekatan tersebut menunjukkan performa yang menjanjikan, sebagian besar penelitian masih menerapkan skema *binary classification* yang hanya menghasilkan label "curang" dan "tidak curang". Pendekatan *binary classification* ini memiliki keterbatasan dalam konteks ujian daring yang kompleks. Pertama, model tidak mampu membedakan jenis atau bentuk kecurangan yang dilakukan, sehingga berbagai perilaku seperti melihat catatan, membuka laman internet, atau menggunakan perangkat lain diperlakukan sebagai satu kategori yang sama (Erdem dan Karabatak, 2025; Henderson dkk., 2023). Kedua, keterbatasan ini menghambat analisis perilaku secara lebih mendalam karena sistem tidak menyediakan informasi detail mengenai pola dan tingkat pelanggaran yang terjadi. Akibatnya, keluaran sistem menjadi kurang informatif dan belum optimal untuk mendukung pengambilan keputusan oleh pengawas maupun institusi pendidikan, seperti penentuan sanksi, evaluasi kebijakan pengawasan, maupun strategi pencegahan kecurangan yang lebih tepat sasaran.

Sebagai respons terhadap keterbatasan tersebut, beberapa penelitian mulai mengembangkan arsitektur yang lebih kompleks, seperti *framework* berbasis *spatio-temporal graph* untuk mengintegrasikan fitur postur tubuh, latar belakang, dan atribut wajah guna meningkatkan akurasi deteksi (Liu, 2024). Pemanfaatan model pralatih (*transfer learning*) juga telah menjadi strategi penting untuk meningkatkan performa pada dataset terbatas, meskipun memerlukan strategi

fine-tuning yang tepat agar tidak terjadi *overfitting* (Becherer dkk., 2019). ResNet50 merupakan salah satu arsitektur CNN modern yang banyak dimanfaatkan sebagai *feature extractor* pada tugas pengenalan citra dan video. ResNet50 menggunakan mekanisme *residual learning* melalui *skip connection* yang membantu menjaga stabilitas pelatihan pada jaringan yang dalam serta mengurangi permasalahan *vanishing gradient* (Sharma, 2022). Keunggulan tersebut menjadikan ResNet50 sebagai salah satu model yang relevan untuk digunakan dalam sistem deteksi perilaku pada lingkungan pembelajaran daring.

Berawal dari perkembangan tersebut, penelitian ini mengembangkan model deteksi kecurangan pada ujian daring berbasis *computer vision* menggunakan kombinasi ResNet50 sebagai *feature extractor* dan BiLSTM sebagai model klasifikasi temporal. Model ini dirancang untuk memproses rekaman video dari kamera *wearable* berbentuk kacamata (*wearcam*) yang menangkap sudut pandang pengguna secara langsung, sehingga aktivitas seperti membaca buku atau penggunaan perangkat lain dapat teridentifikasi secara visual (Thomas dkk., 2022). Klasifikasi dilakukan dalam enam kelas, yaitu satu kelas tidak curang dan lima kelas perilaku kecurangan. Pendekatan ini memungkinkan model tidak hanya mendeteksi keberadaan kecurangan, tetapi juga mengidentifikasi jenis perilaku kecurangan yang dilakukan, sehingga berpotensi memberikan informasi yang lebih spesifik dan bermanfaat bagi pengawas maupun institusi pendidikan pada pengembangan lebih lanjut. Seluruh alur penelitian ditunjukkan pada Gambar 1.

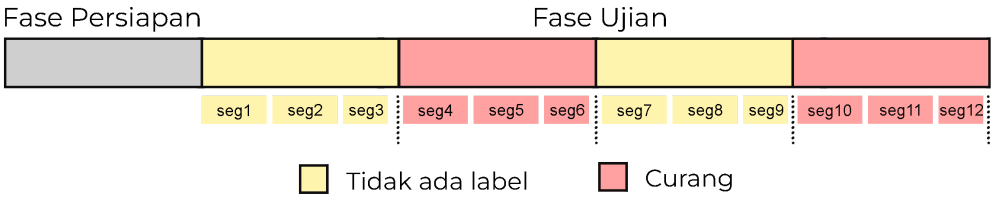


Gambar 1: Diagram Proses Rancangan Penelitian

2 Metode

2.1 Dataset

Penelitian ini menggunakan *Online Exam Proctoring* (OEP) Dataset yang dikembangkan oleh *Computer Vision Laboratory*, *Michigan State University* (MSU) (Atoum dkk., 2017). Dataset ini berisi rekaman audio dan video dari 24 peserta, yang terdiri atas 15 peserta yang mensimulasikan berbagai perilaku kecurangan dan 9 peserta yang mengikuti ujian dengan skenario kecurangan yang dipicu oleh pengawas. Dataset menyediakan beberapa kanal rekaman, yaitu video kamera depan, video *wearable camera* (*wearcam*), dan audio. Penelitian ini menggunakan data video *wearcam* sebagai modalitas utama karena mampu merepresentasikan aktivitas peserta dan lingkungan sekitar secara lebih jelas, seperti membaca buku atau menggunakan ponsel. Sudut pandang *wearcam* yang mengikuti perspektif pengguna secara langsung memberikan informasi visual yang lebih kaya dibandingkan kamera statis, sehingga lebih efektif dalam menangkap berbagai jenis perilaku kecurangan. Dataset OEP hanya menyediakan anotasi untuk lima jenis perilaku kecurangan. Penelitian ini menambahkan label tidak curang melalui anotasi manual pada segmen video yang tidak termasuk dalam *ground truth* kecurangan dan tidak menunjukkan indikasi perilaku mencurigakan berdasarkan pemeriksaan visual. Proses anotasi ditunjukkan pada Gambar 2. Distribusi data video hasil anotasi ditampilkan pada Tabel 1.



Gambar 2: Proses anotasi label tidak curang pada data video

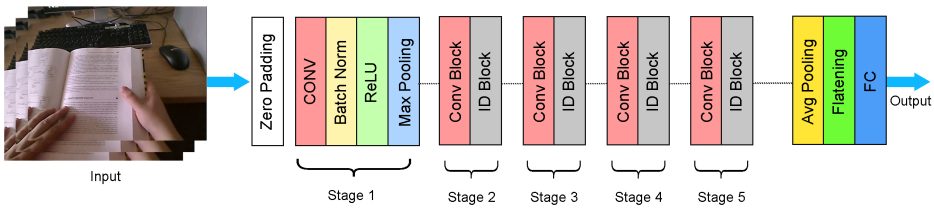
Tabel 1: Distribusi data video pada setiap label

Label	**Jumlah**
Tidak curang	988
Membaca buku	192
Berbicara	285
Menggunakan internet	52
Menelpon	15
Menggunakan ponsel	20

2.2 Pemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan bahwa data video memiliki format dan struktur yang konsisten sebelum memasuki tahap ekstraksi fitur dan klasifikasi. Seluruh video terlebih dahulu dikelompokkan berdasarkan kelasnya, kemudian sistem membaca setiap video dan memeriksa jumlah *frame* yang tersedia. Video yang memiliki jumlah *frame* kurang dari 25 *frame* dieliminasi untuk menjaga konsistensi antar sampel serta mengurangi potensi ketidakseimbangan data. Video yang memenuhi kriteria kemudian diambil sebanyak 25 *frame* yang tersebar secara merata dari awal hingga akhir durasi rekaman. Pendekatan ini memungkinkan setiap video memiliki representasi visual yang proporsional terhadap keseluruhan aktivitas tanpa perlu memproses seluruh *frame*, sehingga dapat meningkatkan efisiensi komputasi (Shen dkk., 2025). Setiap *frame* yang terpilih kemudian diubah ukurannya menjadi 224×224 piksel agar seluruh citra memiliki dimensi yang seragam. Hasil dari tahap prapemrosesan ini berupa struktur data berdimensi $(n, 25, 224, 224, 3)$, dengan n menyatakan jumlah video yang digunakan dalam dataset.

2.3 Ekstraksi Fitur Data Video Menggunakan ResNet50



Gambar 3: Arsitektur ResNet50

ResNet50 menggunakan mekanisme *residual learning* melalui *skip connection* untuk meningkatkan stabilitas pelatihan jaringan yang dalam serta mengurangi permasalahan *vanishing gradient*, sehingga efektif dalam mengekstraksi fitur visual tingkat tinggi dari citra (Sharma, 2022). Arsitektur ResNet50 terdiri atas beberapa *residual stage* yang dimulai dari lapisan konvolusi awal hingga *fully connected layer*, sebagaimana ditunjukkan pada Gambar 3. Secara matematis, jika x merupakan masukan suatu blok residual dan $F(x)$ adalah fungsi transformasi nonlinier yang terdiri atas operasi konvolusi, *batch normalization*, dan fungsi aktivasi, maka keluaran blok residual diformulasikan pada persamaan (1),

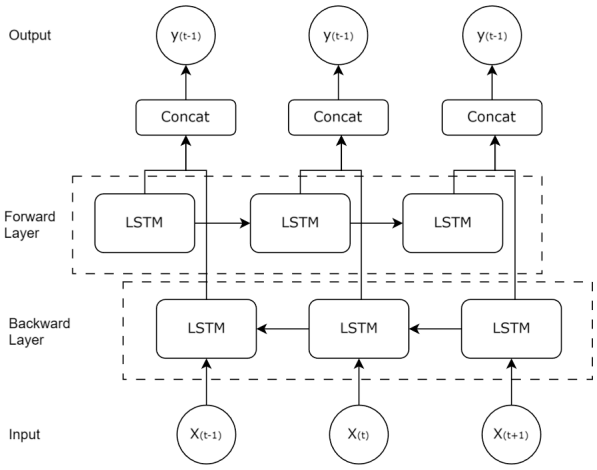
$$H(x) = F(x) + x, \tag{1}$$

dengan $H(x)$ menyatakan keluaran blok residual yang merupakan hasil penjumlahan antara $F(x)$ dan x. Penjumlahan tersebut membentuk *identity mapping* yang memungkinkan jaringan mempelajari fungsi residual secara langsung sehingga mempermudah proses optimisasi pada jaringan yang sangat dalam. Pada proses *backpropagation*, jika L menyatakan fungsi *loss*, maka gradien terhadap masukan x diformulasikan pada persamaan (2),

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial H}\left(\frac{\partial F(x)}{\partial x} + 1\right), \tag{2}$$

dengan suku $+1$ berasal dari jalur identitas sehingga gradien tetap dapat mengalir ke lapisan sebelumnya meskipun nilai $\frac{\partial F(x)}{\partial x}$ sangat kecil. Melalui mekanisme tersebut, ResNet50 mampu mengurangi degradasi kinerja pada jaringan yang dalam dan efektif digunakan sebagai model ekstraksi fitur pada tugas klasifikasi video berbasis *deep learning*.

2.4 *Klasifikasi Menggunakan Bidirectional Long Short-Term Memory* (BiLSTM)



Gambar 4: Arsitektur BiLSTM

Bidirectional Long Short-Term Memory (BiLSTM) merupakan pengembangan dari LSTM yang memproses data sekuensial dalam dua arah, yaitu *forward* dan *backward*. Arsitektur BiLSTM ditunjukkan pada Gambar 4. Berbeda dengan LSTM konvensional yang hanya memproses

urutan data dalam satu arah, BiLSTM menggabungkan keluaran dari kedua arah untuk membentuk representasi fitur akhir yang diformulasikan pada persamaan (3),

$$y(t) = [y_f(t),\, y_b(t)]\,, \tag{3}$$

dengan $y_f(t)$ menyatakan keluaran dari arah *forward* dan $y_b(t)$ menyatakan keluaran dari arah *backward* pada waktu t, sehingga $y(t)$ merupakan gabungan keduanya. Dengan mekanisme tersebut, informasi dari masa lalu maupun masa depan dapat dimanfaatkan secara simultan. Pendekatan ini menghasilkan representasi temporal yang lebih komprehensif dan berpotensi meningkatkan kinerja klasifikasi pada data video yang memiliki dinamika waktu yang kompleks (Liu, 2024). Pada penelitian ini, arsitektur BiLSTM terdiri atas dua lapisan BiLSTM dengan fungsi aktivasi *tanh* yang menghasilkan keluaran berdimensi $n \times 25 \times 512$ dan $n \times 256$. Representasi fitur yang diperoleh selanjutnya diproses melalui lapisan *Dense* dengan fungsi aktivasi *ReLU* berukuran $n \times 128$, kemudian dilanjutkan dengan *Batch Normalization* dan *Dropout* untuk meningkatkan stabilitas pelatihan serta mengurangi risiko *overfitting*. Pada lapisan keluaran, digunakan lapisan *Dense* dengan fungsi aktivasi *Softmax* yang menghasilkan keluaran berdimensi $n \times n_{\text{class}}$ untuk klasifikasi enam kelas.

3 Hasil dan Pembahasan

Penelitian ini mengevaluasi tiga arsitektur *feature extractor*, yaitu ResNet50, ResNet50V2, dan InceptionV3, yang dikombinasikan dengan arsitektur BiLSTM untuk tugas klasifikasi enam kelas pada data video *wearcam*. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *sensitivity*, *specificity*, dan *F1-score*. Distribusi kelas yang digunakan pada skenario klasifikasi enam kelas ditampilkan pada Tabel 1. Perbandingan performa ketiga *feature extractor* pada skenario klasifikasi enam kelas disajikan pada Tabel 2. Hasil eksperimen menunjukkan bahwa kombinasi ResNet50-BiLSTM memberikan performa terbaik dengan *accuracy* sebesar 0.93, rata-rata *F1-score* sebesar 0.85, rata-rata *specificity* sebesar 0.98, dan rata-rata *sensitivity* sebesar 0.80. Hasil ini menunjukkan bahwa fitur yang diekstraksi oleh ResNet50 mampu merepresentasikan pola visual aktivitas kecurangan secara efektif. Struktur *residual connection* pada ResNet50 memungkinkan proses ekstraksi fitur yang lebih stabil pada jaringan yang dalam sehingga informasi penting terkait gerakan dan posisi peserta ujian dapat dipertahankan dengan baik. Dengan memanfaatkan BiLSTM sebagai *classifier*, model dapat menangkap informasi temporal dari urutan *frame* video sehingga pola perilaku peserta ujian dapat dikenali dengan lebih baik. Kombinasi ResNet50V2-BiLSTM menghasilkan *accuracy* sebesar 0.83 dengan rata-rata *F1-score* sebesar 0.62. Meskipun masih mampu mengklasifikasikan sebagian besar kelas dengan cukup baik, performanya mengalami penurunan dibandingkan ResNet50-BiLSTM. Penurunan tersebut terlihat pada hampir seluruh kelas, terutama pada kelas *menggunakan internet* yang mengalami penurunan *F1-score* dari 1.00 menjadi 0.60.

Kombinasi model ResNet50V2-BiLSTM gagal mengenali kelas *menggunakan ponsel*, yang ditunjukkan oleh nilai *F1-score* dan *sensitivity* sebesar 0.00. Nilai 0.00 pada kedua metrik tersebut menunjukkan bahwa model tidak berhasil memprediksi satu pun data uji pada kelas tersebut secara benar (*true positive* = 0), sehingga *sensitivity* bernilai nol dan *F1-score* turut bernilai nol. Kegagalan ini diduga terkait dengan jumlah data pada kelas *menggunakan ponsel* yang relatif sedikit dibandingkan kelas lain (20 video, sebagaimana ditunjukkan pada Tabel 1), sehingga model kurang memperoleh variasi pola visual yang cukup untuk mengenali kelas tersebut pada tahap pengujian. Hasil ini menunjukkan bahwa fitur yang dihasilkan oleh ResNet50V2 kurang

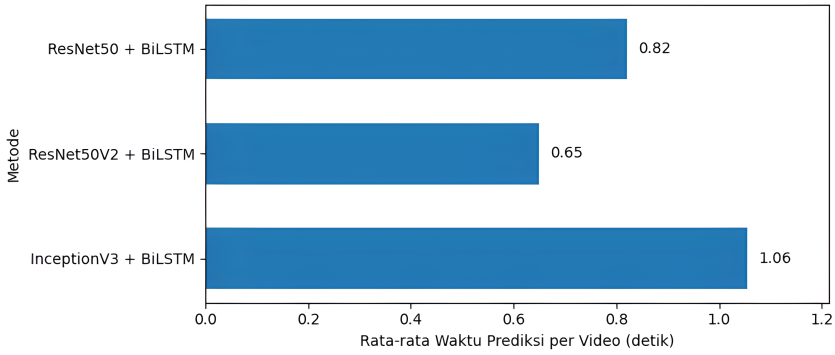
mampu membedakan beberapa aktivitas yang memiliki karakteristik visual yang mirip. Kombinasi InceptionV3-BiLSTM menghasilkan performa terendah dengan *accuracy* sebesar 0.33 dan rata-rata *F1-score* sebesar 0.28. Penurunan performa terjadi hampir pada seluruh kelas jika dibandingkan dengan ResNet50 maupun ResNet50V2. Model ini tidak berhasil mengenali kelas *menelpon* dan kelas *menggunakan ponsel*, yang ditunjukkan oleh nilai *F1-score* sebesar 0.00 pada kedua kelas tersebut. Serupa dengan kegagalan pada ResNet50V2-BiLSTM, nilai 0.00 pada kedua kelas ini menandakan model tidak menghasilkan satu pun prediksi benar pada data uji kelas tersebut. Kegagalan ini konsisten dengan jumlah data pelatihan pada kedua kelas yang tergolong minoritas (masing-masing 15 dan 20 video), sehingga rentan menghasilkan *sensitivity* nol ketika kemampuan ekstraksi fitur model juga kurang optimal, sebagaimana terlihat pada arsitektur InceptionV3. Nilai *F1-score* pada kelas *tidak curang*, *berbicara*, dan *menggunakan internet* juga jauh lebih rendah dibandingkan dua arsitektur lainnya. Hasil ini menunjukkan bahwa pendekatan ekstraksi fitur multiskala pada InceptionV3 kurang sesuai untuk mendeteksi aktivitas kecurangan yang lebih banyak ditandai oleh perubahan gerakan dan posisi peserta ujian dibandingkan variasi objek yang kompleks.

Tabel 2: Perbandingan performa klasifikasi multikelas

Metode	Kelas	F1-Score	Specificity	Sensitivity	Accuracy
ResNet50 + BiLSTM	Tidak curang	0.96	0.93	0.96	0.93
	Membaca buku	0.92	0.99	0.95	
	Berbicara	0.85	0.96	0.86	
	Menggunakan internet	1.00	1.00	1.00	
	Menelpon	0.67	1.00	0.50	
	Menggunakan ponsel	0.67	1.00	0.50	
	Avg	**0.85**	**0.98**	**0.80**	
ResNet50V2 + BiLSTM	Tidak curang	0.89	0.79	0.90	0.83
	Membaca buku	0.86	0.99	0.79	
	Berbicara	0.73	0.93	0.76	
	Menggunakan internet	0.60	0.99	0.60	
	Menelpon	0.67	1.00	0.50	
	Menggunakan ponsel	0.00	0.99	0.00	
	Avg	**0.62**	**0.95**	**0.59**	
InceptionV3 + BiLSTM	Tidak curang	0.43	0.98	0.27	0.33
	Membaca buku	0.62	0.94	0.63	
	Berbicara	0.31	0.93	0.24	
	Menggunakan internet	0.37	0.89	1.00	
	Menelpon	0.00	0.96	0.00	
	Menggunakan ponsel	0.00	0.58	0.00	
	Avg	**0.28**	**0.88**	**0.35**	

Jika dibandingkan pada tingkat kelas, ResNet50 menunjukkan performa yang paling stabil dibandingkan dua arsitektur lainnya. ResNet50 mampu mempertahankan nilai *F1-score* di atas 0.67 pada seluruh kelas dan mencapai nilai sempurna pada kelas *menggunakan internet*. Sebaliknya, ResNet50V2 mengalami kegagalan pada kelas *menggunakan ponsel*, sedangkan InceptionV3 gagal mengenali kelas *menelpon* dan kelas *menggunakan ponsel*. Hasil tersebut menunjukkan bahwa fitur yang dihasilkan ResNet50 lebih mampu membedakan aktivitas yang memiliki kemiripan pola gerakan maupun posisi peserta ujian, termasuk pada kelas-kelas dengan jumlah data yang relatif sedikit. Oleh karena itu, perbedaan performa yang cukup signifikan antar arsitektur

menunjukkan bahwa kualitas representasi fitur memiliki pengaruh yang besar terhadap keberhasilan klasifikasi aktivitas pada data video *wearcam*, terutama pada kondisi data yang tidak seimbang antarkelas.



Gambar 5: Rata-rata waktu prediksi data uji

Dari sisi efisiensi komputasi, hasil pada Gambar 5 menunjukkan adanya perbedaan waktu prediksi pada tiga konfigurasi model yang menggunakan arsitektur BiLSTM dengan ekstraksi fitur yang berbeda. Kombinasi ResNet50V2-BiLSTM menghasilkan waktu prediksi tercepat, yaitu sekitar 0.65 detik per data. ResNet50-BiLSTM memerlukan waktu prediksi sekitar 0.82 detik per data, sedangkan InceptionV3-BiLSTM memiliki waktu prediksi paling tinggi, yaitu sekitar 1.06 detik per data. Perbedaan waktu prediksi tersebut menunjukkan bahwa kompleksitas ekstraksi fitur pada masing-masing arsitektur CNN memengaruhi efisiensi inferensi model. ResNet50V2-BiLSTM mampu menghasilkan prediksi lebih cepat dibandingkan dua konfigurasi lainnya. InceptionV3-BiLSTM membutuhkan waktu komputasi yang lebih besar karena proses ekstraksi fitur yang lebih kompleks. Hasil evaluasi performa klasifikasi menunjukkan bahwa ResNet50-BiLSTM memperoleh nilai yang lebih baik dibandingkan konfigurasi lainnya, khususnya pada kelas-kelas minoritas seperti *menelpon* dan *menggunakan ponsel*. Dengan mempertimbangkan keseimbangan antara performa dan efisiensi, ResNet50-BiLSTM dipilih sebagai konfigurasi terbaik untuk model deteksi kecurangan ujian daring pada penelitian ini.

4 Kesimpulan

Penelitian ini mengembangkan model deteksi kecurangan ujian daring berbasis *deep learning* menggunakan data video *wearcam* dari OEP Dataset. Fitur visual diekstraksi menggunakan ResNet50 yang memanfaatkan mekanisme *residual learning* untuk menghasilkan representasi fitur tingkat tinggi dari setiap *frame* video. Fitur yang diekstraksi selanjutnya dimodelkan secara temporal menggunakan BiLSTM untuk menangkap dependensi waktu dari dua arah. Dari seluruh kombinasi yang dievaluasi, ResNet50-BiLSTM memberikan performa terbaik pada klasifikasi enam kelas dengan *accuracy* sebesar 0,93, rata-rata *F1-score* sebesar 0,85, rata-rata *specificity* sebesar 0,98, dan rata-rata *sensitivity* sebesar 0,80. Pemilihan *feature extractor* terbukti memberikan pengaruh yang lebih besar terhadap performa klasifikasi dibandingkan variasi arsitektur RNN, dengan ResNet50 secara konsisten unggul dibandingkan ResNet50V2 dan InceptionV3 pada seluruh konfigurasi. Secara keseluruhan, kombinasi ResNet50 dan BiLSTM terbukti efektif dalam mendeteksi dan mengklasifikasikan berbagai jenis kecurangan pada ujian daring. Penelitian selanjutnya dapat difokuskan pada peningkatan keragaman data untuk mengatasi keti-

dakseimbangan kelas, khususnya pada kelas-kelas minoritas seperti menelpon orang lain dan menggunakan ponsel, serta eksplorasi strategi augmentasi data dan arsitektur yang lebih efisien untuk meningkatkan kemampuan generalisasi model.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi S1 Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

Pustaka

Ababneh, K. I., Ahmed, K., dan Dedousis, E. (2022). Predictors of cheating in online exams among business students during the covid pandemic: Testing the theory of planned behavior. *International Journal of Management Education*, 20.

Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D., dan Liu, X. (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia*, 19.

Becherer, N., Pecarina, J., Nykl, S., dan Hopkinson, K. (2019). Improving optimization of convolutional neural networks through parameter fine-tuning. *Neural Computing and Applications*, 31.

Child, F., Frank, M., Law, J., dan Sarakatsannis, J. (2023). What do higher education students want from online learning? Technical report.

Erdem, B. dan Karabatak, M. (2025). Cheating detection in online exams using deep learning and machine learning. *Applied Sciences*, 15.

Essahraui, S., Lamaakal, I., Maleh, Y., Makkaoui, K. E., Bouami, M. F., Ouahbi, I., Almousa, M., AlQahtani, A. A. S., dan El-Latif, A. A. A. (2025). Deep learning models for detecting cheating in online exams. *Computers, Materials & Continua*, 85:3151–3183.

Fendler, R., Beard, D., dan Godbey, J. (2024). A robust examination of cheating on unproctored online exams. *Electronic Journal of e-Learning*, 22.

Henderson, M., Chung, J., Awdry, R., Ashford, C., Bryant, M., Mundy, M., dan Ryan, K. (2023). The temptation to cheat in online exams: moving beyond the binary discourse of cheating and not cheating. *International Journal for Educational Integrity*, 19:21.

Kaddoura, S. dan Gumaei, A. (2022). Towards effective and efficient online exam systems using deep learning-based cheating detection approach. *Intelligent Systems with Applications*, 16.

Lamba, S. dan Sharma, D. N. (2024). Learning-based multimodal cheating detection in online proctored exams. *Journal of Electrical System*.

Lindín, C., Engel, A., Gràcia, M., Rivera-Vargas, P., dan Rubio, M. J. (2023). Literature review on emerging educational practices mediated by digital technologies in higher education, based on academic papers.

Liu, C. (2024). Long short-term memory (lstm)-based news classification model. *PLoS ONE*, 19.

Newton, P. M. dan Essex, K. (2024). How common is cheating in online exams and did it increase during the covid-19 pandemic? a systematic review.

Sakhipov, A., Omirzak, I., dan Fedenko, A. (2025). Beyond face recognition: A multi-layered approach to academic integrity in online exams. *Electronic Journal of e-Learning*, 23.

Sharma, S. (2022). Analyzing effect on residual learning by gradual narrowing fully-connected

layer width and implementing inception block in convolution layer. *Journal of Computer Science*, 18.

Shen, G., Sun, Y.-H., Hu, Y. H., dan Jiang, H. (2025). Sampling strategies for efficient training of deep learning object detection algorithms.

Smirani, L. K. dan Boulahia, J. A. (2022). An algorithm based on convolutional neural networks to manage online exams via learning management system without using a webcam. *International Journal of Advanced Computer Science and Applications*, 13.

Thomas, G., Bennie, J. A., Cocker, K. D., Andriyani, F. D., Booker, B., dan Biddle, S. J. (2022). Using wearable cameras to categorize the type and context of screen-based behaviors among adolescents: Observational study. *JMIR Pediatrics and Parenting*, 5.

Turan, Z., Kucuk, S., dan Karabey, S. C. (2022). The university students' self-regulated effort, flexibility and satisfaction in distance education. *International Journal of Educational Technology in Higher Education*, 19.