JURNAL RISET TEKNIK KOMPUTER
Halaman Jurnal: https://journal.smartpublisher.id/index.php/jurtikom
Halaman UTAMA Jurnal : https://journal.smartpublisher.id

DOI: https://doi.org/10.69714/38wan661

IMPLEMENTASI DETEKSI KEKERASAN DENGAN PERINGATAN VISUAL MENGGUNAKAN *MOBILENETV2* DAN *BIDERECTIONAL LONG SHORT-TERM MEMORY (Bi-LSTM)*

Larasati [a*], Wahyu S.J. Saputra [b]
[a] Fakultas Ilmu Komputer, Sains Data, 22083010018@student.upnjatim.ac.id, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Kota Surabaya, Jawa Timur
[b] Fakultas Ilmu Komputer, Sains Data, wahyu.s.j.saputra.if@upnjatim.ac.id, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Kota Surabaya, Jawa Timur
*Korespondensi

ABSTRACT

This study developed an automatic video-based violence detection system by combining the MobileNetV2 architecture and Bidirectional Long Short-Term Memory (BiLSTM). MobileNetV2 is used to efficiently extract spatial features from each video frame, while BiLSTM is utilized to understand the temporal relationships between frames so that violence patterns can be recognized more accurately. The model is designed to classify videos into two classes: "Violence" and "Non-Violence." The training process showed a consistent increase in accuracy with a decrease in loss values, without any indication of overfitting. The resulting model achieved an accuracy of 92%, with high precision and recall values, especially in detecting violent actions. The system is implemented in real-time using input from a live camera. When violence is detected, the system automatically displays a visual warning in the form of a red screen with the text "VIOLENCE DETECTED!" and saves a frame clip with a timestamp as documentation. Testing shows that the system can distinguish violent contexts effectively without producing significant false positives. With reliable performance and quick response times, this system has great potential for application in CCTV surveillance in public areas, schools, and conflict-prone regions as an efficient, proactive, and adaptive AI-based solution.

Keywords: *Violence Detection, MobileNetV2, BiLSTM, Real-Time, Artificial Intelligence*

ABSTRAK

Penelitian ini mengembangkan sistem deteksi kekerasan otomatis berbasis video dengan menggabungkan arsitektur *MobileNetV2* dan *Bidirectional Long Short-Term Memory (BiLSTM). MobileNetV2* digunakan untuk mengekstraksi fitur spasial dari setiap frame video secara efisien, sementara *BiLSTM* dimanfaatkan untuk memahami hubungan temporal antar frame sehingga pola kekerasan dapat dikenali secara lebih akurat. Model dirancang untuk mengklasifikasikan video ke dalam dua kelas, yaitu "*Violence*" dan "*Non-Violence*". Proses pelatihan menunjukkan peningkatan akurasi secara konsisten dengan penurunan nilai *loss*, tanpa indikasi *overfitting*. Model yang dihasilkan mencapai akurasi sebesar 92%, dengan nilai *precision* dan *recall* yang tinggi, terutama dalam mendeteksi aksi kekerasan. Sistem ini diimplementasikan secara *real-time* menggunakan input dari kamera langsung. Ketika kekerasan terdeteksi, sistem secara otomatis menampilkan peringatan visual berupa layar merah dengan teks "KEKERASAN TERDETEKSI!" serta menyimpan cuplikan frame beserta *timestamp* sebagai dokumentasi. Pengujian menunjukkan bahwa sistem mampu membedakan konteks kekerasan dengan baik tanpa menghasilkan *false positive* yang signifikan. Dengan performa yang handal dan respons cepat, sistem ini berpotensi besar untuk diterapkan dalam pengawasan *CCTV* di area publik, sekolah, dan wilayah rawan konflik sebagai solusi berbasis kecerdasan buatan yang efisien, proaktif, dan adaptif.

Kata Kunci: Deteksi Kekerasan, *MobileNetV2*, *BiLSTM*, *Real-Time*, Kecerdasan Buatan

Received April 8, 2025; Revised April 11, 2025; Accepted Mei 28, 2025; Published Juni 3, 2025

1. PENDAHULUAN

Kekerasan di ruang publik menjadi fenomena sosial yang kian mengkhawatirkan di berbagai belahan dunia, termasuk Indonesia. Masyarakat yang semakin padat, interaksi sosial yang kompleks, serta lemahnya pengawasan menjadi faktor yang memperburuk kondisi tersebut. Berdasarkan laporan UNICEF berjudul *Hidden in Plain Sight*, sekitar 40% remaja usia 13 - 15 tahun pernah mengalami kekerasan fisik, dan 50% pernah mengalami perundungan di sekolah dalam satu tahun terakhir (Unicef, 2015). Kekerasan tidak hanya terjadi di lingkungan kriminal, tetapi juga muncul di ruang terbuka seperti stasiun, sekolah, pusat perbelanjaan, bahkan transportasi publik. Sementara itu, laporan dari Kementerian Pemberdayaan Perempuan dan Perlindungan Anak menunjukkan bahwa dari Januari hingga awal 2024, terdapat 1.318 kasus kekerasan fisik (Abdillah et al., n.d.). Keberadaan *CCTV* yang melimpah nyatanya belum cukup mampu mencegah atau mendeteksi kekerasan secara tepat waktu. Hal ini disebabkan karena sebagian besar sistem *CCTV* masih bergantung pada pemantauan manual oleh operator yang terbatas daya perhatian dan responsnya. Situasi ini berisiko tinggi menyebabkan kekerasan luput dari pantauan dan penanganan lambat. Dalam penelitian (Negre et al., 2024) dijelaskan bahwa deteksi kekerasan secara otomatis menjadi solusi jangka pendek paling efisien untuk mencegah kehilangan nyawa akibat keterlambatan respons. Kamera pengawas seharusnya tidak hanya menjadi alat dokumentasi pasca-kejadian, tetapi juga menjadi perangkat aktif yang mampu mendeteksi dan memberikan peringatan sejak awal. Oleh sebab itu, pengembangan sistem otomatis untuk mendeteksi kekerasan dalam video secara *real-time* menjadi sangat krusial. Deteksi otomatis akan membantu mengidentifikasi kekerasan sejak dini dan memberikan peringatan sebelum insiden memburuk.

Kemajuan teknologi kecerdasan buatan (*AI*) dan *deep learning* membuka peluang besar untuk menyelesaikan persoalan ini. Sistem cerdas mampu belajar dari kumpulan data video kekerasan dan mengenali pola visual serta perilaku mencurigakan. Salah satu pendekatan yang umum digunakan adalah kombinasi *Convolutional Neural Network (CNN)* untuk ekstraksi spasial dan *Long Short-Term Memory (LSTM)* untuk pemodelan temporal. Namun, penggunaan arsitektur berat seperti *DenseNet 3D* atau *Transformer* seringkali menghambat performa *real-time* karena memerlukan daya komputasi tinggi (Rendón-Segador et al., 2021) Oleh karena itu, diperlukan arsitektur ringan yang tetap akurat, terutama jika ingin diterapkan pada perangkat terbatas seperti sistem *CCTV edge-based*. Dalam hal ini, MobileNet dikenal sebagai *CNN* ringan yang sangat efisien, sedangkan *Bi-LSTM* menawarkan keunggulan dalam memahami konteks urutan dua arah. Penggunaan *MobileNetV2* dan *Bi-LSTM* diyakini mampu menjaga keseimbangan antara akurasi dan efisiensi sistem. Selain itu, pendekatan ini sangat cocok diterapkan pada sistem yang membutuhkan respons cepat, seperti keamanan publik berbasis video. Penerapan *deep learning* tidak hanya terbatas pada deteksi objek, tetapi kini berkembang pada analisis tindakan manusia yang kompleks. Termasuk di antaranya adalah tindakan kekerasan yang seringkali berlangsung sangat cepat dan tidak selalu mencolok secara visual.

Beberapa penelitian sebelumnya telah mengembangkan sistem deteksi kekerasan berbasis *deep learning*, namun belum seluruhnya menjawab tantangan efisiensi dan implementasi nyata. Penelitian oleh (Dzakwanul Fikri et al., 2017) menggunakan arsitektur *3D CNN* untuk deteksi kekerasan pada *CCTV* berbasis *Jetson Nano* dan mencatat akurasi sebesar 90%, tetapi sistem belum optimal dari sisi kecepatan respon dan hanya efektif dalam jarak pendek (Dzakwanul Fikri et al., 2017) Di sisi lain, penelitian oleh (Khalfaoui et al., n.d.) menggabungkan *MobileNetV3* dan *LSTM* dengan perhatian temporal dua arah (BiLTMA), dan terbukti akurat pada empat dataset berbeda (Khalfaoui et al., n.d.). Namun, arsitektur mereka masih cukup kompleks untuk diterapkan di perangkat *edge* dengan keterbatasan memori. Penelitian (Mane et al., 2024) membuktikan bahwa integrasi *MobileNet* dan *Bi-LSTM* dapat mendeteksi berbagai jenis anomali secara *real-time* dengan akurasi hingga 95,33% (Mane et al., 2024) Meski demikian, pendekatan tersebut masih fokus pada deteksi umum (*anomaly*) dan belum spesifik terhadap kekerasan sebagai objek utama. Dari sisi pendekatan *machine learning* tradisional, penelitian oleh (Supartini et al., 2022) lebih menekankan deteksi kekerasan pada remaja melalui edukasi dan pendampingan, bukan berbasis *video surveillance* (Supartini et al., 2022). Ini menunjukkan perlunya sistem berbasis visual yang mampu bekerja tanpa intervensi manusia dalam mendeteksi kekerasan di ruang publik. (Patel, n.d.) dalam *Real-Time Violence Detection Using CNN-LSTM* menyebutkan bahwa peringatan visual yang sederhana namun *real-time* dapat meningkatkan efektivitas sistem pengawasan. Namun, sistem tersebut hanya menggunakan *LSTM* dan tidak menggabungkan arsitektur *CNN* ringan seperti *MobileNet*, sehingga waktu inferensinya lebih lama.

Dalam skala sistem komprehensif, *CMS* yang diperkenalkan oleh (Mukto et al., 2024) dalam Intelligent Systems with Applications mencakup deteksi senjata, kekerasan, dan wajah. Walaupun sistem ini cukup

lengkap, pendekatannya terlalu umum dan tidak fokus pada pendeteksian kekerasan secara mendalam. Studi oleh (Zhang et al., 2022) dalam penelitiannya membuktikan bahwa arsitektur ringan *MobileNet-TSM* mampu mendeteksi kekerasan dengan akurasi tinggi namun tidak menggunakan unit rekuren seperti *LSTM*, sehingga kurang memahami hubungan antarframe yang kompleks. Dalam penelitian oleh (Mohod, 2024) disebutkan bahwa pemahaman konteks temporal jangka pendek dan panjang sangat krusial dalam deteksi kekerasan. Pendekatan CNN konvensional maupun model *Decision Tree* tidak mampu mengenali pola kekerasan halus yang terjadi secara dinamis. Hal ini menjadi dasar penting untuk memilih arsitektur yang mampu merekam informasi spasial dan temporal secara bersamaan. Salah satu kelemahan pendekatan lain seperti *DenseNet 3D* adalah tingginya jumlah parameter dan kebutuhan memori yang besar (Rendón-Segador et al., 2021). Sementara pendekatan *Transformer* yang digunakan oleh (Soeleman et al., 2022) juga memiliki keterbatasan saat diterapkan pada video berkualitas rendah dan data terbatas. (Silva Deena et al., 2022) dalam penelitiannya masih menggunakan metode *machine learning* klasik yang kurang optimal dalam mengenali kekerasan kompleks di lingkungan nyata. Oleh karena itu, pengembangan sistem berbasis *MobileNetV2* dan *Bi-LSTM* memberikan solusi yang lebih ringan, fleksibel, serta mampu mengatasi permasalahan video surveillance di dunia nyata.

Penelitian ini mengambil posisi berbeda dengan mengusulkan sistem deteksi kekerasan berbasis *MobileNetV2* dan *Bi-LSTM* yang dilengkapi dengan fitur interaktif berupa peringatan visual *real-time*. Sistem akan menampilkan teks "kekerasan terdeteksi" serta kedipan layar jika aksi kekerasan dikenali. Pendekatan ini tidak hanya memberikan klasifikasi pasif, tetapi langsung terhubung ke mekanisme notifikasi sebagai bentuk respons cepat. Penggabungan *CNN* ringan dan unit rekuren dua arah memungkinkan sistem untuk memahami dinamika gerakan kekerasan secara lebih mendalam. Selain itu, dibandingkan sistem seperti *ViolenceNet* yang hanya fokus pada akurasi klasifikasi tanpa memperhatikan efisiensi implementasi, sistem ini lebih adaptif terhadap kebutuhan di lapangan. Sistem juga dirancang agar mudah diintegrasikan dengan kamera *CCTV* yang sudah ada, tanpa perlu perangkat keras tambahan berspesifikasi tinggi. Dengan demikian, pendekatan ini tidak hanya kuat secara teknis, tetapi juga relevan dalam penerapannya di dunia nyata. Sistem ini ditujukan untuk ruang publik seperti stasiun, sekolah, atau mall, di mana kecepatan deteksi dan interaksi visual sangat penting.

Selain mengutamakan performa, penelitian ini juga menekankan pada efisiensi sistem. Pemilihan *MobileNetV2* sebagai *backbone CNN* didasarkan pada kebutuhan akan pemrosesan cepat di perangkat terbatas. Sementara itu, *Bi-LSTM* memungkinkan pemodelan informasi temporal dua arah, yang sangat penting dalam mendeteksi kekerasan yang berkembang dalam hitungan detik. Sistem ini juga dirancang agar dapat diperluas ke perangkat *IoT* atau *edge device*, membuka kemungkinan integrasi dengan sistem keamanan pintar berbasis *Internet of Things*. Dengan demikian, sistem yang dikembangkan tidak hanya sebagai alat pendeteksi kekerasan, tetapi juga menjadi bagian dari solusi keamanan yang lebih luas. Pendekatan yang diusulkan mampu menjembatani kebutuhan dunia nyata dengan teknologi *deep learning* yang mutakhir. Penggabungan efisiensi, ketepatan, dan interaktivitas menjadi kekuatan utama sistem ini dalam menjawab tantangan keamanan publik berbasis *video surveillance*.

Dengan mempertimbangkan berbagai penelitian terdahulu dan tantangan teknis yang ada, penelitian ini bertujuan untuk mengembangkan sistem deteksi kekerasan otomatis berbasis video dengan integrasi fitur peringatan visual secara *real-time*. Sistem ini menggunakan kombinasi *MobileNetV2* sebagai ekstraktor fitur spasial dan *Bi-LSTM* sebagai pemodel hubungan temporal dua arah. Dibandingkan pendekatan lain, sistem ini menonjol dalam hal efisiensi model, kecepatan respon, dan kemampuan adaptasi terhadap kualitas video yang buruk. Fitur peringatan visual menjadi elemen baru yang memperkuat respons langsung terhadap insiden kekerasan. Dengan pendekatan ini, sistem diharapkan mampu meningkatkan kualitas pemantauan keamanan di ruang publik secara signifikan. Penelitian ini merupakan langkah awal menuju sistem keamanan berbasis *AI* yang ringan, cepat, dan responsif. Sistem ini juga dirancang agar mudah diterapkan dan terintegrasi dengan ekosistem keamanan digital masa depan.

2. METODOLOGI PENELITIAN

Penelitian ini membangun sistem deteksi kekerasan secara *real-time* dalam video dengan pendekatan *deep learning.* Tujuan dari sistem ini adalah untuk secara otomatis mengenali aksi kekerasan dalam video pengawasan dan memberikan peringatan visual secara langsung kepada pengguna. Untuk mencapai tujuan tersebut, digunakan kombinasi arsitektur *MobileNetV2* sebagai ekstraktor fitur spasial dari frame video, dan *Bidirectional Long Short-Term Memory (BiLSTM)* sebagai pemodel temporal untuk memahami urutan gerakan antar frame. Pendekatan ini dipilih karena efisien secara komputasi dan memiliki kemampuan yang kuat dalam mengenali pola aktivitas yang bersifat sekuensial, seperti kekerasan dalam video. Penelitian

terdahulu juga menunjukkan bahwa kombinasi *CNN* ringan dan *LSTM* dua arah mampu memberikan hasil akurat dengan latensi rendah, sangat cocok untuk sistem berbasis *edge computing* (Mane et al., 2024). Proses pengembangan sistem dilakukan secara bertahap, dimulai dari pengambilan dataset, ekstraksi data, pelatihan model, hingga implementasi sistem secara *real-time*.

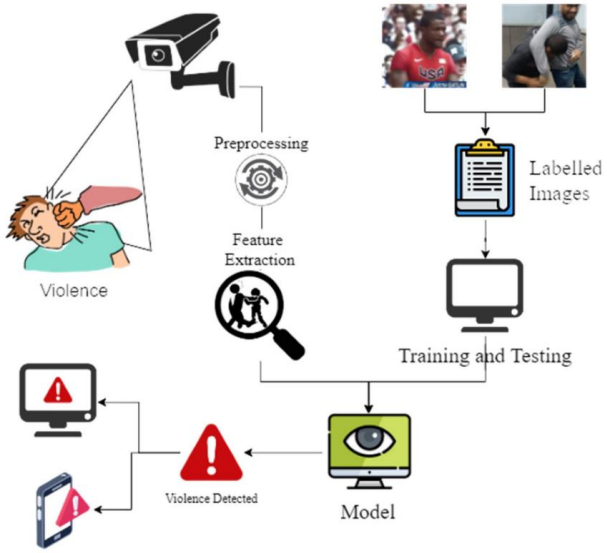
2.1 Dataset

Dataset yang digunakan pada penelitian ini adalah *Real-Life Violence Situations* Dataset yang diambil dari Kaggle. Dataset ini terdiri dari dua kelas utama, yaitu:

a. *Violence*: video-video pendek yang berisi aksi kekerasan seperti perkelahian, pukulan, atau aksi agresif lainnya.
b. *NonViolence*: video-video tanpa kekerasan seperti orang berjalan, berbicara, atau duduk.

Setiap video berdurasi antara 1 hingga 5 detik, dengan resolusi 720p. Jumlah total video pada dataset ini adalah 1000 klip kekerasan dan 1000 klip non-kekerasan. Dataset ini dipilih karena sudah digunakan dalam berbagai penelitian sebelumnya dan representatif untuk simulasi lingkungan *CCTV*.

2.2 Alur Penelitian



Gambar 1. Alur penelitian

Alur penelitian ini melibatkan beberapa tahap utama yang saling berurutan, dimulai dari proses pengumpulan data, ekstraksi fitur, pelatihan model, hingga tahap implementasi sistem deteksi secara *real-time*. Alur ini dirancang agar seluruh tahapan dapat berjalan secara terintegrasi, mulai dari input video hingga sistem dapat memberikan peringatan ketika kekerasan terdeteksi. Setiap video diproses menjadi urutan frame, diekstraksi fitur visualnya menggunakan *MobileNetV2*, dan urutan fitur tersebut dianalisis oleh *BiLSTM* untuk menentukan apakah mengandung kekerasan atau tidak. Jika kekerasan terdeteksi, sistem akan langsung memberikan peringatan visual kepada pengguna. Alur ini menggambarkan pendekatan *end-to-end* dalam penerapan visi komputer berbasis *deep learning* untuk mendeteksi aktivitas tertentu dalam video.

2.3 Ekstraksi Frame

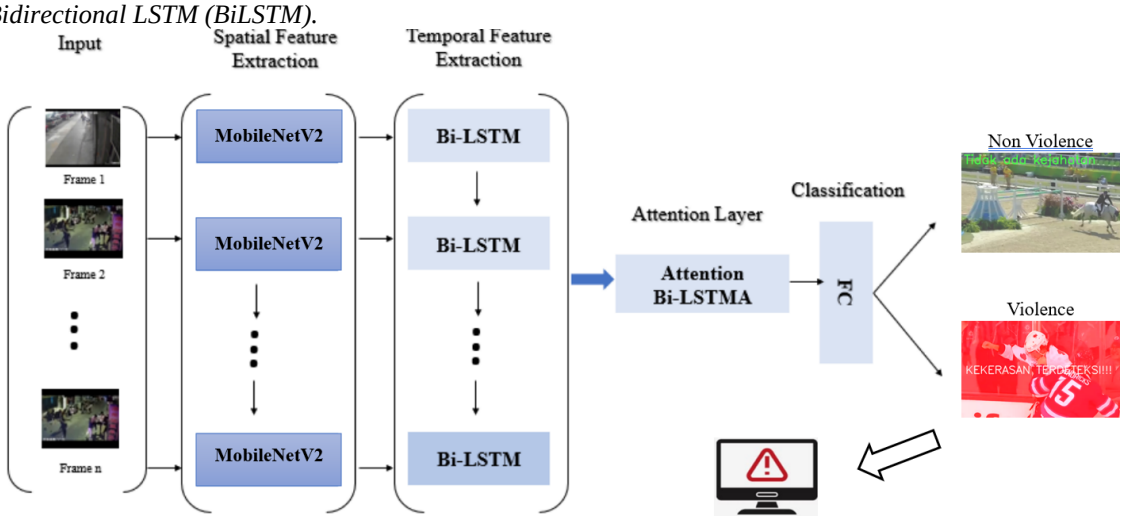
Tahapan ekstraksi frame bertujuan untuk mengubah video menjadi data gambar yang dapat diproses oleh model *deep learning*. Dari setiap video dalam dataset, diambil sebanyak 20 frame secara merata berdasarkan panjang durasi video tersebut. Frame diambil dengan interval waktu yang tetap agar representasi gerakan dalam video dapat terwakili dengan baik. Setiap frame kemudian diubah ukurannya menjadi 224×224 piksel agar sesuai dengan input standar *MobileNetV2*, lalu dilakukan normalisasi piksel ke dalam rentang nilai 0 hingga 1 untuk memudahkan proses pelatihan model. Hasil dari proses ini adalah array berdimensi (20, 224, 224, 3) untuk setiap video. Frame-frame ini kemudian disusun sebagai urutan (sequence) dan disimpan ke dalam list fitur (*features*), sedangkan label video (0 untuk *NonViolence* dan 1 untuk *Violence*) disimpan ke dalam list label (*labels*).

2.4 Ektraksi feature

Setelah frame diekstraksi dan dinormalisasi, setiap frame diproses menggunakan *MobileNetV2* untuk mengambil representasi fitur visual. *MobileNetV2* merupakan *CNN* ringan yang memiliki performa baik dalam pengenalan objek pada perangkat dengan sumber daya terbatas, namun tetap memiliki kemampuan ekstraksi fitur yang kuat (Mane et al., 2024). Model ini telah dilatih sebelumnya (*pre-trained*) pada dataset *ImageNet* dan digunakan dalam kondisi tanpa lapisan klasifikasi akhir (*include_top=False*). Dengan demikian, *MobileNetV2* hanya digunakan untuk menghasilkan representasi spasial dari gambar, berupa vektor fitur berdimensi rendah namun informatif. Ekstraksi ini dilakukan secara *time-distributed*, artinya setiap frame dalam urutan dianalisis secara individu namun dalam konteks sekuensial. Output dari tahap ini berupa urutan fitur yang kemudian akan digunakan sebagai input untuk model *BiLSTM*.

2.5 Arsitektur Model

Model deteksi kekerasan dalam penelitian ini terdiri dari dua komponen utama, yaitu *MobileNetV2* dan *Bidirectional LSTM (BiLSTM)*.



Gambar 2. Model penelitian

a. *MobileNetV2*

MobileNetV2 memiliki keunggulan sebagai model *lightweight CNN* yang cepat dan tidak memerlukan sumber daya komputasi tinggi, namun tetap mampu mengekstraksi fitur penting dari gambar. *MobileNetV2* bertugas mengekstraksi fitur dari setiap frame secara individual. Digunakan sebagai ekstraktor fitur dari masing-masing frame video. Model ini sudah dilatih sebelumnya (*pre-trained*) pada dataset *ImageNet* dan digunakan tanpa lapisan output akhir (*include_top=False*). Output dari *MobileNetV2* adalah fitur spasial berdimensi rendah namun informatif. *MobileNetV2* telah terbukti unggul dalam sistem *real-time* video *surveillance* karena performanya yang seimbang antara akurasi dan kecepatan (Khalfaoui et al., n.d.).

b. *Bidirectional Long Short-Term Memory (BiLSTM)*

BiLSTM digunakan untuk memodelkan hubungan temporal antarframe. Keunggulan *BiLSTM* dibanding *LSTM* biasa adalah kemampuannya membaca sekuens dari dua arah sekaligus, sehingga konteks kekerasan yang terjadi di tengah atau akhir urutan frame tetap dapat dikenali (Dzakwanul Fikri et al., 2017). Hasil dari *BiLSTM* diteruskan ke *layer Dense* dengan aktivasi *Softmax* untuk mengklasifikasikan dua kelas (*Violence* dan *NonViolence*). Model dilatih menggunakan *categorical_crossentropy* dan *optimizer SGD*, serta didukung dengan *EarlyStopping* dan *ReduceLROnPlateau* guna mencegah *overfitting*.

2.6 Pelatihan dan Validasi Model

Proses pelatihan model dilakukan dengan membagi dataset menjadi dua bagian, yaitu 80% data untuk pelatihan dan 20% untuk validasi. Data pelatihan digunakan untuk membangun model, sedangkan data validasi digunakan untuk mengevaluasi generalisasi model terhadap data yang belum pernah dilihat. Selama proses pelatihan, metrik yang digunakan untuk memantau kinerja model antara lain *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Evaluasi ini penting untuk memastikan bahwa model tidak hanya mampu mengenali kelas mayoritas, tetapi juga seimbang dalam mendeteksi kedua kelas. Pendekatan evaluasi beragam ini umum digunakan dalam pengujian sistem berbasis *machine learning* untuk deteksi kekerasan (Abdillah et al., n.d.). Model yang menunjukkan performa terbaik pada data

validasi disimpan dalam format .h5 untuk digunakan kembali pada tahap implementasi. Pemilihan metrik yang beragam juga memberikan gambaran menyeluruh terhadap kekuatan dan kelemahan model dalam mendeteksi kekerasan dalam berbagai kondisi.

2.7 Implementasi *Real-Time Detection*

Model yang telah dilatih diintegrasikan ke dalam sistem *real-time* dengan alur berikut:

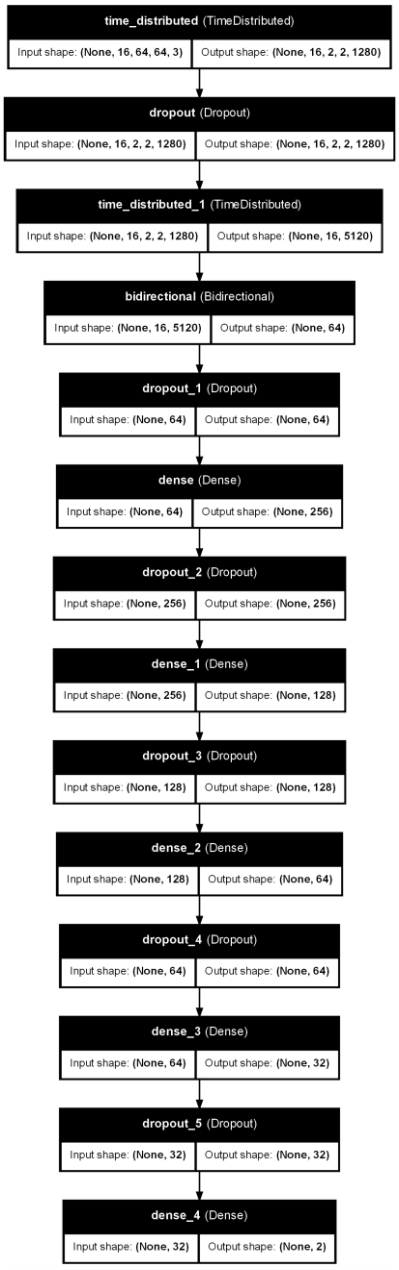
a. Input video berasal dari *webcam* atau video file (.mp4) yang dibaca menggunakan *cv2.VideoCapture*.
b. Setiap 30 frame dikumpulkan sebagai satu sequence untuk dianalisis.
c. Setiap frame diresize, dinormalisasi, dan diekstrak fiturnya menggunakan *MobileNetV2*.
d. Fitur sequence dikirim ke model *BiLSTM* untuk prediksi apakah kekerasan terjadi atau tidak.
e. Jika model memprediksi *Violence*, maka sistem secara otomatis:
    1) Menampilkan layar berkedip merah dengan *overlay* teks “KEKERASAN TERDETEKSI!”
    2) Menyimpan *timestamp* dan cuplikan frame ke dalam folder bukti.

Output ini ditampilkan melalui jendela *OpenCV* secara *real-time* dan sistem akan terus berjalan hingga proses dihentikan oleh pengguna. Dengan pendekatan ini, sistem mampu memberikan sinyal dini terhadap insiden kekerasan tanpa perlu identifikasi wajah atau pelaku, sehingga proses lebih ringan dan responsif.

3. HASIL DAN PEMBAHASAN

3.1 Arsitektur Model

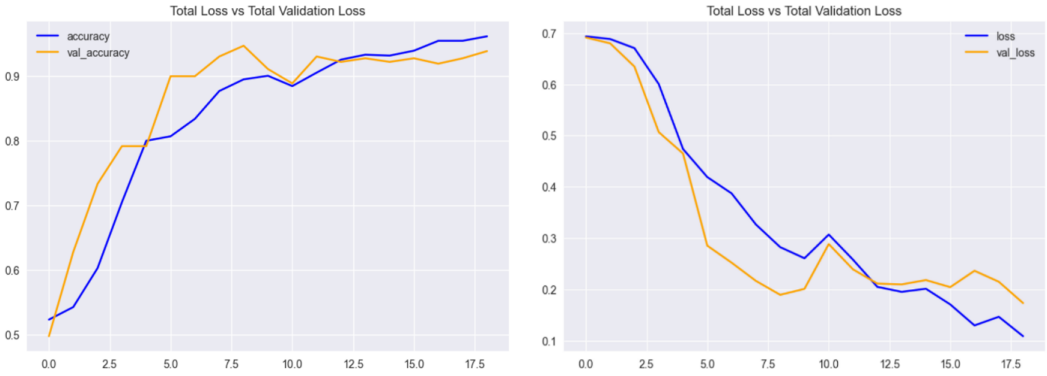
Model yang digunakan adalah gabungan *MobileNetV2* dan *BiLSTM*. *MobileNetV2* berfungsi sebagai ekstraktor fitur spasial dari frame video. Setiap frame video diproses menggunakan *TimeDistributed* yang membungkus *MobileNetV2*. Hasil ekstraksi fitur dari semua frame kemudian dirangkum secara temporal oleh lapisan *BiLSTM*. Lapisan *BiLSTM* digunakan untuk menangkap hubungan antarwaktu (urutan frame). Setelah itu, output *BiLSTM* masuk ke beberapa lapisan *Dense* untuk proses klasifikasi. Beberapa lapisan *Dropout* disisipkan untuk mencegah *overfitting*. Output akhir model berupa dua kelas yaitu "*Violence*" dan "*NonViolence*". Arsitektur ini dirancang untuk memahami baik informasi spasial maupun temporal dari video input.



Gambar 3. Arsitektur model

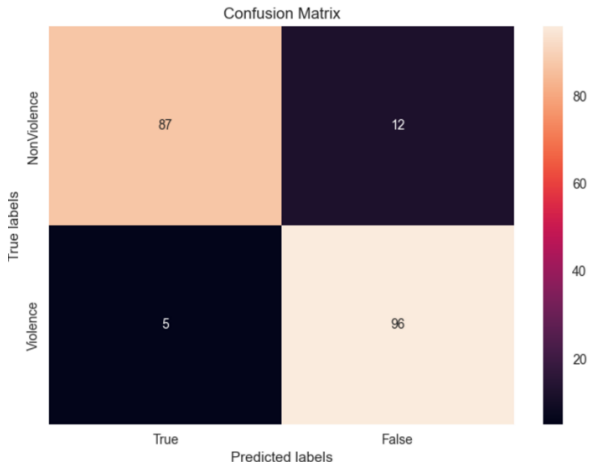
3.2 Evaluasi Model

Evaluasi performa model dilakukan menggunakan beberapa metrik utama, yaitu *accuracy*, *loss*, *confusion matrix*, serta *precision*, *recall*, dan *F1-score*. Grafik akurasi menunjukkan tren peningkatan yang konsisten seiring bertambahnya *epoch* pelatihan, dengan akurasi validasi yang mencapai > 90%. Hal ini mengindikasikan bahwa model memiliki kemampuan belajar yang stabil dan mampu menangkap pola dari data. Grafik loss yang terus menurun dan tidak menunjukkan selisih mencolok antara nilai loss pada data latih dan validasi menjadi indikator kuat bahwa model tidak mengalami *overfitting*.



Gambar 4. Grafik akurasi dan loss

Hasil *confusion matrix* memperlihatkan model berhasil mengklasifikasikan sebagian besar data dengan benar. Terdapat 87 data *non-violence* yang diklasifikasikan benar, dan 96 data *violence* yang juga tepat. Meski ada kesalahan prediksi, jumlahnya relatif kecil, yaitu 12 untuk *non-violence* dan 5 untuk *violence*



Gambar 5. Confusion matrix

Pada *classification report*, model mencapai akurasi keseluruhan sebesar 92%, menunjukkan performa yang cukup tinggi. *Precision* untuk kelas *non-violence* mencapai 0.95, sedangkan kelas *violence* 0.89. *Recall* kelas *violence* lebih tinggi (0.95) dibandingkan *non-violence* (0.88), artinya model lebih sensitif terhadap kekerasan.

	precision	recall	f1-score	support
non-violence	0.95	0.88	0.91	99
violence	0.89	0.95	0.92	101
accuracy			0.92	200
macro avg	0.92	0.91	0.91	200
weighted avg	0.92	0.92	0.91	200

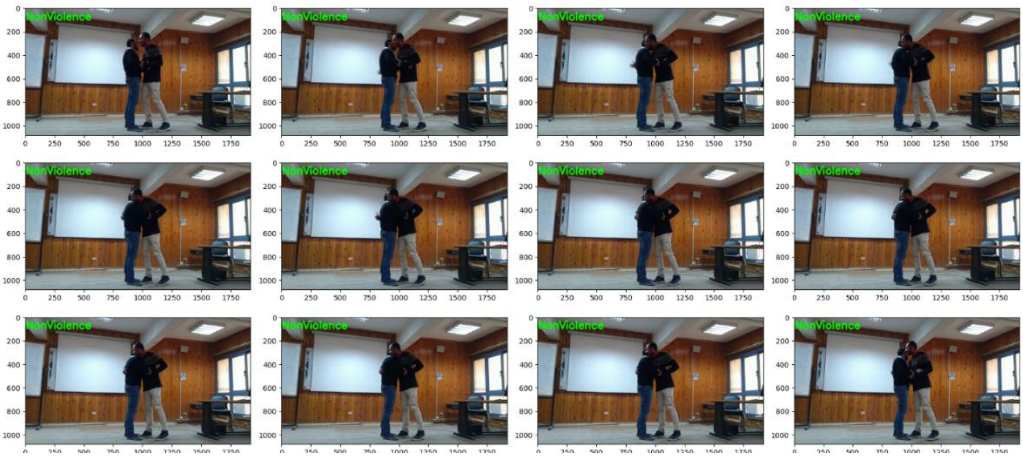
Gambar 6 Classification report

Evaluasi ini membuktikan bahwa model mampu mengenali kekerasan dalam video dengan baik. Kombinasi *MobileNetV2* dan *BiLSTM* terbukti efektif untuk menangani data spasial dan temporal secara bersamaan.

3.3 Prediksi

Berdasarkan hasil validasi pada beberapa cuplikan video, model berhasil mengklasifikasikan kategori *violence* dan *non-violence* dengan baik. Pada video yang mengandung aksi kekerasan, model secara konsisten memberikan label "*Violence*" pada sebagian besar frame. Sementara itu, pada video tanpa kekerasan, seluruh frame berhasil dikenali sebagai "*NonViolence*" tanpa kesalahan deteksi. Hal ini menunjukkan bahwa model mampu membedakan konteks kekerasan secara akurat dan stabil, baik secara spasial maupun temporal. Secara keseluruhan, hasil ini menguatkan bahwa model yang dikembangkan efektif dan layak digunakan dalam sistem deteksi kekerasan otomatis berbasis video.

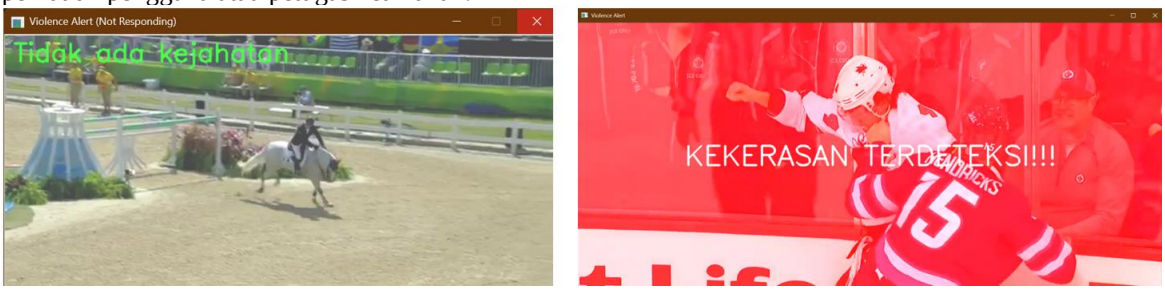
Gambar 7. Prediksi Violence



Gambar 8. Prediksi NonViolence

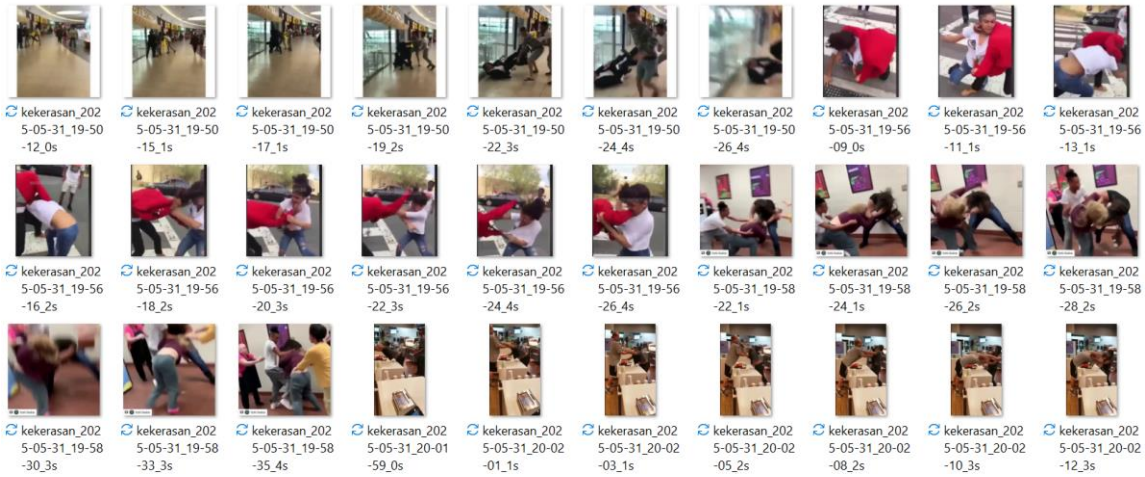
3.4 Implementasi

Sistem bekerja dengan cara memproses setiap frame video yang ditangkap dari kamera secara langsung, lalu menganalisisnya untuk mendeteksi adanya indikasi kekerasan. Jika model memprediksi suatu frame atau serangkaian frame mengandung kekerasan (label "*Violence*"), maka sistem akan secara otomatis memicu respons visual berupa layar berkedip merah serta menampilkan teks *overlay* besar bertuliskan "Kekerasan Terdeteksi!" untuk memberi peringatan yang jelas dan instan. Efek kedipan ini dibuat dengan mengubah warna latar belakang atau menambahkan *layer* transparan merah pada tampilan, untuk menarik perhatian pengguna atau petugas keamanan.



Gambar 9. Implementasi

Gambar di sisi kanan memperlihatkan bagaimana tampilan sistem ketika kekerasan terdeteksi layar menjadi merah dan teks "KEKERASAN TERDETEKSI!!!" muncul secara mencolok. Sementara gambar di kiri menunjukkan situasi non-kekerasan, seperti pertandingan olahraga, yang secara tepat tidak ditandai oleh sistem sebagai kekerasan membuktikan bahwa sistem memiliki kemampuan membedakan konteks secara akurat dan tidak mudah menghasilkan *false positive*.



Gambar 10. Cuplikan frame dokumentasi dalam folder

Selain memberi peringatan visual, sistem ini juga memiliki fitur dokumentasi otomatis. Ketika kekerasan terdeteksi, sistem secara langsung menyimpan cuplikan frame yang terdeteksi beserta timestamp (waktu kejadiannya) ke dalam folder khusus sebagai bukti visual. Hal ini penting untuk keperluan pelaporan, verifikasi, atau tindakan lanjut yang mungkin diperlukan oleh pihak berwenang.

Implementasi seperti ini sangat potensial digunakan pada sistem *CCTV* publik, sekolah, stasiun, atau area rawan konflik, karena tidak hanya mendeteksi tetapi juga mendokumentasikan kejadian penting secara otomatis dan efisien. Dengan pendekatan ini, sistem tidak hanya menjadi alat deteksi, tetapi juga alat pencegahan dan penegakan keamanan berbasis teknologi *AI*.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun sistem deteksi kekerasan otomatis berbasis video menggunakan kombinasi arsitektur *MobileNetV2* dan *Bidirectional LSTM (BiLSTM). MobileNetV2* berperan sebagai ekstraktor fitur spasial yang ringan namun efektif, sementara *BiLSTM* mampu memahami pola temporal dari urutan frame video. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 92%, dengan *precision* dan *recall* yang tinggi pada kedua kelas, terutama dalam mengenali aksi kekerasan. Model mampu bekerja secara stabil dan tidak mengalami *overfitting*, dibuktikan dari tren grafik akurasi dan loss yang konsisten selama pelatihan. Implementasi *real-time* sistem ini mampu memberikan peringatan visual yang responsif ketika kekerasan terdeteksi. Pendekatan ini menunjukkan bahwa teknologi *deep learning* dapat digunakan secara efisien dalam sistem pengawasan berbasis video.

Selain mendeteksi kekerasan, sistem juga dilengkapi dengan fitur dokumentasi otomatis berupa penyimpanan cuplikan frame dan waktu kejadian. Fitur ini meningkatkan potensi penggunaan sistem dalam lingkungan nyata seperti *CCTV* publik, sekolah, atau area rawan konflik. Sistem ini terbukti akurat dalam klasifikasi, dan juga mampu menghindari *false positive* dengan baik. Proses deteksi dilakukan tanpa memerlukan identifikasi wajah atau pelaku, sehingga lebih ringan dan privasi tetap terjaga. Dengan performa tinggi dan efisiensi komputasi yang baik, sistem ini layak dijadikan solusi praktis untuk mendeteksi dan mencegah kekerasan secara proaktif. Ke depannya, sistem ini dapat dikembangkan lebih lanjut dengan integrasi ke perangkat edge dan kamera pintar untuk cakupan pemantauan yang lebih luas.

DAFTAR PUSTAKA

[1] F. Abdillah, A. Khoiriyah, A. N. Aziz, and I. G. W. Politeknik Negeri Jember, “Sistem Deteksi Kekerasan Real-Time menggunakan YOLOv5 untuk Keamanan Publik,” *SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, n.d. [Online]. Available: https://doi.org/10.31284/p.snestik.2024.5861

[2] A. D. Fikri, F. Utaminingrum, and E. G. E. Setyawan, “Sistem Pendeteksi Kekerasan di Ruang Publik Menggunakan Metode 3D Convolutional Neural Network,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[3] A. Khalfaoui, A. Badri, and I. El Mourabit, “Efficient Violence Detection with Bi-directional Motion Attention and MobileNetV3-LSTM,” n.d. [Online]. Available: https://ssrn.com/abstract=5120792

[4] D. Mane, S. Phatangare, S. Nawale, S. Wani, V. Gujarathi, and V. Loya, "Real-Time Anomaly Detection in Video Surveillance: A Mathematical Modeling and Nonlinear Analysis Perspective with MobileNet and Bi-LSTM," *Communications on Applied Nonlinear Analysis*, vol. 31, no. 2s, 2024.
[5] N. P. Mohod, "Real-Time Violence Detection in Surveillance Videos Using Deep Learning Approach," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 4, pp. 1267–1274, 2024. [Online]. Available: https://doi.org/10.22214/ijraset.2024.59968
[6] M. M. Mukto, M. Hasan, M. M. Al Mahmud, I. Haque, M. A. Ahmed, T. Jabid, M. S. Ali, M. R. A. Rashid, M. M. Islam, and M. Islam, "Design of a real-time crime monitoring system using deep learning techniques," *Intelligent Systems with Applications*, vol. 21, 2024. [Online]. Available: https://doi.org/10.1016/j.iswa.2023.200311
[7] P. Negre, R. S. Alonso, A. González-Briones, J. Prieto, and S. Rodríguez-González, "Literature Review of Deep-Learning-Based Detection of Violence in Video," *Sensors*, vol. 24, no. 12, 2024. [Online]. Available: https://doi.org/10.3390/s24124016
[8] M. B. Patel, "Real-Time Violence Detection Using CNN-LSTM," n.d.
[9] F. J. Rendón-Segador, J. A. Álvarez-García, F. Enríquez, and O. Deniz, "Violencenet: Dense multi-head self-attention with bidirectional convolutional LSTM for detecting violence," *Electronics (Switzerland)*, vol. 10, no. 13, 2021. [Online]. Available: https://doi.org/10.3390/electronics10131601
[10] J. Silva Deena, M. T. Ahammed, U. M. Boppana, M. Afroj, S. Ghosh, S. Hossain, and P. Balaji, "Real-time based Violence Detection from CCTV Camera using Machine Learning Method," in *2022 International Conference on Industry 4.0 Technology, I4Tech 2022*, 2022. [Online]. Available: https://doi.org/10.1109/I4Tech55392.2022.9952805
[11] M. A. Soeleman, C. Supriyanto, D. P. Prabowo, and P. N. Andono, "Video Violence Detection Using LSTM and Transformer Networks Through Grid Search-Based Hyperparameters Optimization," *International Journal of Safety and Security Engineering*, vol. 12, no. 05, pp. 615–622, 2022. [Online]. Available: https://doi.org/10.18280/ijsse.120510
[12] Y. Supartini, E. S. Tambunan, T. Suheri, and R. Ningsih, "Pengembangan Model Deteksi Dini Kekerasan pada Remaja sebagai Upaya Meningkatkan Kemampuan dalam Mendeteksi Adanya Kekerasan pada Remaja," *Quality: Jurnal Kesehatan*, vol. 16, no. 2, pp. 82–95, 2022. [Online]. Available: https://doi.org/10.36082/qjk.v16i2.792
[13] UNICEF, *Hidden in Plain Sight: A Statistical Analysis of Violence Against Children*. New York: United Nations Children's Fund, 2014. [Online]. Available: https://www.unicef.org/reports/hidden-plain-sight
[14] Y. Zhang, Y. Li, and S. Guo, "Lightweight mobile network for real-time violence recognition," *PLoS ONE*, vol. 17, no. 10, 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0276939