



## Analisis Perbandingan Performa Model ConvLSTM dan LRCN dalam Pengenalan Aktivitas Gerak Manusia

Amir Hamzah<sup>1\*</sup>, Jamilatul Badriyah<sup>2</sup>

<sup>1</sup>Fakultas Teknik, Prodi Informatika, Universitas Madura, Indonesia

<sup>2</sup>Universitas Madura, Indonesia

Alamat: Jl. Raya Panglegur No.Km 3,5, Barat, Panglegur, Kec. Tlanakan, Kabupaten Pamekasan, Jawa Timur 69371

Korespondensi penulis: [amirzha03@gmail.com](mailto:amirzha03@gmail.com)\*

**Abstract.** *This study compares the performance of two deep learning models, namely Convolutional Long Short-Term Memory (ConvLSTM) and Long-term Recurrent Convolutional Network (LRCN), in the task of recognizing human activity from videos. Human activity recognition is an important field in computer vision with many applications, such as security monitoring, human-computer interaction, and social media-based video analysis. ConvLSTM is a model that combines convolution operations with long-term memory LSTM, thus capable of capturing spatial and temporal information simultaneously. This approach is ideal for processing video data sequences that have spatial and temporal dimensions. On the other hand, LRCN combines the power of spatial feature extraction from Convolutional Neural Network (CNN) and temporal sequence modeling through Recurrent Neural Network (RNN), specifically LSTM, to understand movement patterns in videos. The study used the UCF50 dataset consisting of 50 activity classes, but was limited to five classes for the focus of the experiment. The dataset was divided into 80% for training and 20% for testing, and the model was drilled for 50 epochs using early stopping to prevent overfitting. The results show that both models have high training performance. ConvLSTM achieved a training accuracy of around 98% and a validation accuracy of 90%, while LRCN achieved a training accuracy of 99.5% and a validation accuracy of 88%. Although ConvLSTM demonstrated good stability on the validation data, further testing using TikTok videos as real-world data showed that LRCN had a higher confidence level in recognizing activities, with most predictions achieving confidence scores above 80%. This difference in performance indicates that while ConvLSTM excels in generalizing on training data, LRCN is more robust to real-world data variations.*

**Keywords:** *ConvLSTM, Deep Learning, Human Activity Recognition, LRCN.*

**Abstrak.** Penelitian ini membandingkan performa dua model deep learning, yaitu Convolutional Long Short-Term Memory (ConvLSTM) dan Long-term Recurrent Convolutional Network (LRCN), dalam tugas pengenalan aktivitas manusia dari video. Pengenalan aktivitas manusia merupakan bidang penting dalam computer vision yang memiliki banyak aplikasi, seperti pemantauan keamanan, interaksi manusia-komputer, dan analisis video berbasis media sosial. ConvLSTM merupakan model yang menggabungkan operasi konvolusi dengan memori jangka panjang LSTM, sehingga mampu menangkap informasi spasial dan temporal secara simultan. Pendekatan ini sangat ideal untuk memproses urutan data video yang memiliki dimensi ruang dan waktu. Di sisi lain, LRCN menggabungkan kekuatan ekstraksi fitur spasial dari Convolutional Neural Network (CNN) dan pemodelan urutan temporal melalui Recurrent Neural Network (RNN), khususnya LSTM, untuk memahami pola pergerakan dalam video. Penelitian menggunakan dataset UCF50 yang terdiri dari 50 kelas aktivitas, namun dibatasi pada lima kelas untuk fokus eksperimen. Dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, dan model dilatih selama 50 epoch menggunakan early stopping untuk mencegah overfitting. Hasil menunjukkan bahwa kedua model memiliki performa pelatihan yang tinggi. ConvLSTM mencapai akurasi pelatihan sekitar 98% dan akurasi validasi 90%, sedangkan LRCN mencapai akurasi pelatihan 99,5% dan validasi 88%. Meskipun ConvLSTM menunjukkan kestabilan yang baik pada data validasi, pengujian lebih lanjut menggunakan video TikTok sebagai data dunia nyata menunjukkan bahwa LRCN memiliki tingkat kepercayaan yang lebih tinggi dalam mengenali aktivitas, dengan sebagian besar prediksi mencapai confidence score di atas 80%. Perbedaan performa ini menunjukkan bahwa meskipun ConvLSTM unggul dalam generalisasi data pelatihan, LRCN lebih tangguh dalam menghadapi variasi data di dunia nyata.

**Kata kunci:** ConvLSTM, Deep Learning, LRCN, Pengenalan Aktivitas Gerak Manusia.

## **1. LATAR BELAKANG**

Seiring dengan perkembangan dari Artificial Intelligence dari waktu ke waktu, banyak hal yang dapat membantu manusia untuk mengefisienkan banyak hal. Hal ini terbukti dengan semakin bertambahnya metode-metode terkait teknologi pemrosesan gambar yang salah satunya dapat berfungsi untuk menganalisis aktivitas yang dilakukan oleh manusia (Zhou, Porikli, Crandall, Van Gool, & Wang, 2023). Untuk melakukan analisis terhadap pengenalan aktivitas manusia, terdapat beberapa metode yang dapat digunakan, diantaranya adalah metode ConvLSTM dan LRCN.

Ketika dikombinasikan dengan LSTM, Convolutional Neural Network (CNN) menjadi salah satu model yang dikenal dengan efektif dalam kasus deep learning model (Wu et al., 2024). ConvLSTM adalah varian hibrida dari arsitektur LSTM yang menggunakan operator konvolusi daripada perkalian matriks untuk masukan keadaan dan transisi keadaan ke keadaan (Rahman, Islam, Moon, Tasnim, & Siddique, 2022). Hal ini memungkinkan algoritma untuk memproses data spasial (gambar) dan juga temporal (urutan waktu/frame). Ketika hanya menggunakan CNN, kita dapat mengetahui terdapat objek dalam suatu gambar. Namun ketika menggunakan gabungan dari CNN dan LSTM (ConvLSTM) kita dapat mengetahui aktivitas yang terdapat dalam gambar tersebut berdasarkan urutan frame nya (Komatsu et al., 2021).

Adapun metode yang cocok untuk melakukan pemrosesan gambar adalah metode LRCN. Untuk pemodelan urutan temporal, Long-Term Recurrent Convolutional Network (LRCN) menggabungkan Convolutional Neural Network dengan Long Short-Term Memory (LSTM) (Mansour, Escorcia-Gutierrez, Gamarra, Villanueva, & Leal, 2021). CNN berperan untuk melakukan ekstraksi informasi spasial gambar, sedangkan lapisan LSTM berperan untuk mengirimkan informasi spasial tersebut pada setiap langkah waktu. CNN sangat baik dalam menemukan pola dalam data lokal. Pola ini kemudian digunakan untuk mengelompokkan gambar yang sangat cocok untuk melakukan pengenalan terhadap aktivitas manusia (Ismail, Aziz, Kasim, & Daud, 2021).

Pada artikel ini, penulis ingin memberikan suatu wawasan terkait bagaimana performa dari model ConvLSTM dan LRCN dalam mengenali aktivitas gerak manusia (Almars, 2021). Artikel ini diharapkan mampu untuk memberikan wawasan kepada para peneliti berikutnya terkait model manakah yang lebih cocok digunakan ketika ingin melakukan pengenalan terhadap aktivitas gerak manusia (Elmetwally, Eldeeb, & Elmougy, 2025).

## 2. KAJIAN TEORITIS

### Penelitian Sebelumnya

Agar artikel ini memiliki landasan yang jelas, penulis merujuk terhadap beberapa jurnal terdahulu yang membahas terkait pengenalan aktivitas gerak manusia. Berdasarkan penelitian yang dilakukan oleh (Lina, Augustine, Stephen, & Salim, 2024) yang membahas terkait “Pengenalan Aktivitas Manusia Dalam Ruang Dengan Convolutional Neural Networks”. Pada jurnal tersebut, tujuan utama yang dilakukan adalah bagaimana cara untuk melakukan pengenalan terhadap aktivitas manusia yang dapat dilakukan secara otomatis. Aktivitas pengawasan tersebut dilakukan secara live maupun direkam melalui kamera CCTV yang dapat mengenali aktivitas seperti duduk, berdiri, belajar, mengangkat tangan, dan bertepuk tangan. Agar dapat mengenali seluruh aktivitas tersebut, penulis pada jurnal tersebut menggunakan sistem dari metode Convolution Neural Network dengan menggunakan data uji yang berasal dari data internet maupun rekaman video yang telah dilakukan sebelumnya. Dari hal tersebut, hasil menunjukkan akurasi pengenalan sebesar 86,65% untuk data uji dari internet dan 96% untuk data uji dari dari IP Camera. Tidak hanya itu, seluruh kegiatan yang terjadi di dalam tempat dan waktu tersebut di rangkum dalam sebuah sistem yang dapat merangkum keseluruhan catatan aktivitas.

Agar laporan ini memiliki landasan yang jelas, penulis merujuk terhadap beberapa jurnal terdahulu yang membahas terkait pengenalan aktivitas gerak manusia. Berdasarkan penelitian yang dilakukan oleh (Lina et al., 2024) yang membahas terkait “**Pengenalan Aktivitas Manusia Dalam Ruang Dengan Convolutional Neural Networks**”. Pada jurnal tersebut, tujuan utama yang dilakukan adalah bagaimana cara untuk melakukan pengenalan terhadap aktivitas manusia yang dapat dilakukan secara otomatis. Aktivitas pengawasan tersebut dilakukan secara live maupun direkam melalui kamera CCTV yang dapat mengenali aktivitas seperti duduk, berdiri, belajar, mengangkat tangan, dan bertepuk tangan. Agar dapat mengenali seluruh aktivitas tersebut, penulis pada jurnal tersebut menggunakan sistem dari metode Convolution Neural Network dengan menggunakan data uji yang berasal dari data internet maupun rekaman video yang telah dilakukan sebelumnya. Dari hal tersebut, hasil menunjukkan akurasi pengenalan sebesar 86,65% untuk data uji dari internet dan 96% untuk data uji dari dari IP Camera. Tidak hanya itu, seluruh kegiatan yang terjadi di dalam tempat dan waktu tersebut di rangkum dalam sebuah sistem yang dapat merangkum keseluruhan catatan aktivitas.

## **Python**

Python merupakan bahasa pemrograman yang sering dipakai untuk membangun website, software, machine learning dan deep learning, serta mengotomatisasikan tugas dan melakukan analisis data. Python sendiri merupakan bahasa pemrograman yang berhasil diracik oleh Guido Van Rossum. Bahasa pemrograman ini bisa digunakan untuk membuat berbagai program yang berbeda, tidak khusus untuk masalah tertentu saja. Karena sifatnya yang serbaguna dan mudah untuk digunakan, python menjadi bahasa pemrograman yang paling banyak digunakan, terutama untuk programmer yang masih pemula.

## **ConvLSTM**

ConvLSTM adalah jaringan saraf konvolusional yang digabungkan dengan jaringan LSTM, sehingga mirip dengan LSTM. Namun, operasi konvolusionalnya terjadi selama transisi lapisan. Untuk gambar dan video yang melibatkan ketergantungan waktu, ConvLSTM adalah yang terbaik. Jaringan ini mengumpulkan korelasi lokal-spasial. Menggunakan masukan dan keadaan masa lalu tetangga lokalnya, ConvLSTM menentukan keadaan masa depan sel tertentu dalam grid. Operator konvolusi untuk transisi state-to-state dan input-to-state sangat mudah digunakan (Vrskova, Hudec, Kamencay, & Sykora, 2022).

## **LRCN**

LRCN (Long-term Recurrent Convolutional Networks) adalah arsitektur pembelajaran mendalam yang menggunakan CNN dan RNN untuk memproses data video, terutama untuk pengenalan aktivitas manusia (HAR). Model ini dapat menangani fitur spasial dan temporal secara bersamaan, yang memungkinkan untuk mengenali pola jangka panjang dan kompleks dalam data video. LRCN juga menunjukkan keunggulan dalam mengatasi dependensi temporal jangka panjang dan kompleksitas pola temporal yang mungkin sulit ditangani oleh model lain seperti ConvLSTM. Selain itu, karena mengintegrasikan kedua jenis lapisan, LRCN memiliki tingkat kompleksitas yang lebih tinggi, yang dapat meningkatkan kemungkinan mengenali pola jangka panjang dan kompleks dalam data video (Uddin et al., 2024).

## **Pengenalan Aktivitas Gerak Manusia**

Pengenalan aktivitas gerak manusia atau sering kali dikenal dengan HAR (Human Activity Recognition) merupakan suatu bidang studi yang berfokus pada deteksi otomatis aktivitas harian yang dilakukan oleh orang berdasarkan rekaman deret waktu dari sensor (Arshad, Bilal, & Gani, 2022). Ini dapat berupa tindakan sederhana maupun kompleks, seperti makan, minum, menyikat gigi, atau tindakan yang melibatkan transisi tertentu. HAR menggunakan data dari sensor seperti IMU yang tertanam di tubuh dan dapat dilakukan secara

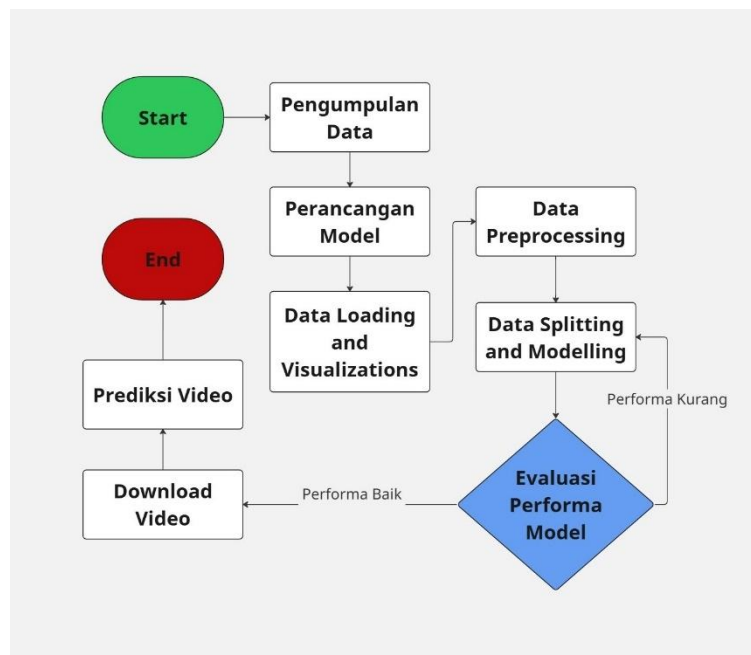
online maupun offline. Tujuan utamanya adalah untuk mengklasifikasi dan memantau aktivitas manusia dalam bidang seperti biometrik, kesehatan, olahraga, dan keamanan (Gupta, 2021).

### 3. METODE PENELITIAN

Dalam menganalisis perbandingan performa dari model ConvLSTM dan LRCN dalam pengenalan aktivitas gerak manusia tentunya diperlukan metode-metode yang digunakan selama penelitian berlangsung. Berikut ini adalah metode-metode yang peneliti lakukan selama penelitian.

#### Flowchart Penelitian

Dalam melakukan suatu penelitian, diperlukan sebuah flowchart agar proses pelaksanaan penelitian berjalan dengan baik dan sesuai dengan alur yang telah ditetapkan. Berikut ini adalah flowchart penelitian yang dapat dilihat pada Gambar 1 berikut.



**Gambar 1.** Flowchart Penelitian

Berdasarkan flowchart yang ada pada Gambar 1 diatas, proses pertama yang dilakukan selama penelitian ini berlangsung adalah dengan melakukan pengumpulan data. Setelah data dikumpulkan, proses selanjutnya yang perlu dilakukan adalah dengan melakukan perancangan model. Dalam merancang suatu model, beberapa tahap yang peneliti lakukan mencakup Data Loading and Visualizations, Data Preprocessing, Data Splitting and Modelling, Evaluasi Performa Model, Download Video, dan Prediksi Video. Tahapan-tahapan tersebut merupakan keseluruhan proses yang dilakukan oleh peneliti selama melakukan penelitian.

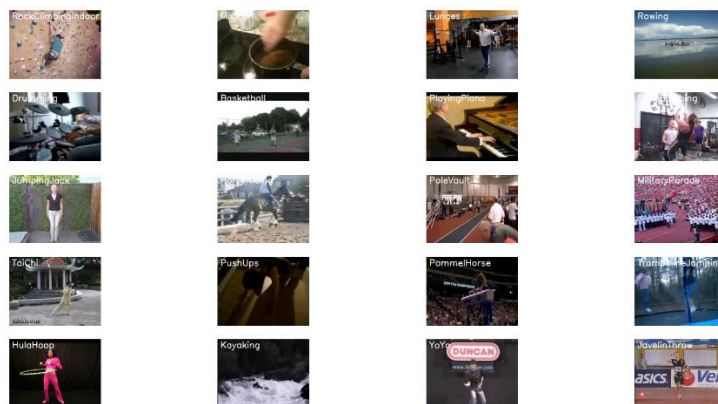
## Pengumpulan Data

Proses pertama yang dilakukan oleh peneliti untuk melakukan pengenalan aktivitas gerak manusia adalah dengan melakukan pengumpulan data terkait aktivitas gerak manusia. Adapun peneliti berhasil mengumpulkan data Realistic Action Recognition UCF50 yang peneliti dapatkan pada situs Kaggle. Data tersebut berjenis video file dengan total kelas sebanyak 50 aktivitas manusia.

## Perancangan Model

### a. Data Loading and Visualizations

Setelah melalui proses pengumpulan data, tahapan selanjutnya adalah dengan melakukan Data Loading and Visualizations. Tahapan ini dilakukan dengan tujuan agar data bisa digunakan untuk dilakukan pemodelan lebih lanjut serta melakukan visualisasi terhadap data yang akan digunakan. Berikut ini adalah hasil dari visualisasi yang dapat ditampilkan pada Gambar 2 dibawah ini.



**Gambar 2.** Visualisasi Data UCF50

Dari Gambar 2 tersebut dapat dilihat bahwa dari total 50 jenis kelas, peneliti hanya menampilkan total 20 jenis sampel yang ada pada data UCF50. Kelas tersebut divisualisasikan dalam bentuk gambar yang telah diberikan label di tiap gambar nya

### b. Data *Preprocessing*

Kemudian setelah melakukan tahapan Data Loading and Visualizations, langkah selanjutnya adalah dengan melakukan Data Preprocessing terhadap data yang akan peneliti gunakan. Pada tahapan ini, peneliti mengatur nilai awal dari image height, image width, sequence length agar dapat diterapkan dengan sama rata terhadap semua jenis video yang akan di proses. Setelah menentukan nilai awal tersebut, peneliti juga menentukan 5 kelas yang akan digunakan selama penelitian berlangsung agar tidak terlalu lama dalam memproses data yang terdapat sebanyak 50 kelas. Beberapa kelas yang telah dipilih adalah PlayingGuitar, YoYo, Billiards, SkateBoarding dan Basketball.

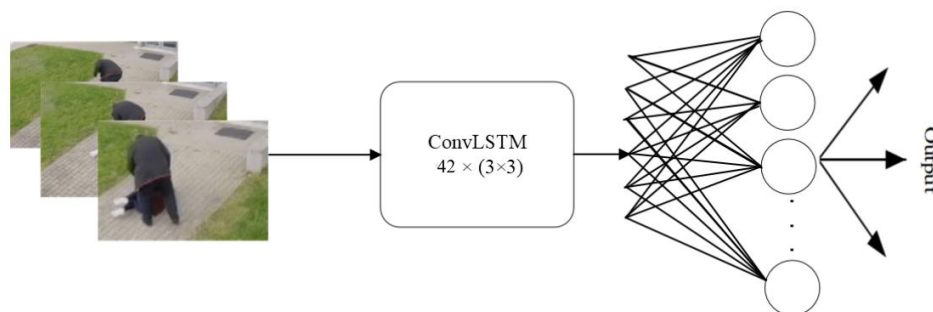
Selanjutnya, peneliti melakukan ekstraksi terhadap frame yang dibutuhkan dari video. Hal ini dilakukan dengan tujuan agar ukuran dari frame dilakukan resizing berdasarkan nilai image height dan image width yang telah ditentukan nilainya. Setelah melakukan resizing terhadap frame, langkah selanjutnya adalah dengan melakukan normalisasi terhadap frame agar berada pada rentang 0 hingga 1.

Adapun setelah melakukan proses ekstraksi frame, peneliti melakukan pembuatan dataset baru berdasarkan features, video file paths dan labels atau kelas yang telah ditentukan sebelumnya. Hal ini dilakukan agar model yang akan digunakan tidak memproses keseluruhan kelas yang telah di download pada data UCF50. Langkah terakhir setelah melakukan pembuatan dataset baru adalah dengan melakukan one hot encoding terhadap label yang digunakan.

### c. Data Splitting and Modelling

Tahapan selanjutnya dalam melakukan analisis pengenalan aktivitas gerak manusia adalah dengan melakukan Data Splitting and Modelling. Data Splitting dilakukan untuk membagi data yang akan di modelkan menjadi data test dan data train. Data tersebut akan dibagi masing-masing kedalam dua bagian, yakni *features\_train* dan *features\_test*, lalu *labels\_train* dan *labels\_test*. Setelah itu, data akan dilatih sebanyak 80% dari data keseluruhan dan 20% untuk data pengujian.

Kemudian, setelah selesai melakukan data splitting, langkah selanjutnya adalah dengan melakukan modelling dengan menggunakan metode yang telah ditentukan sebelumnya, yakni metode ConvLSTM dan metode LRCN. Berikut ini adalah bagaimana model menggunakan metode ConvLSTM bekerja.

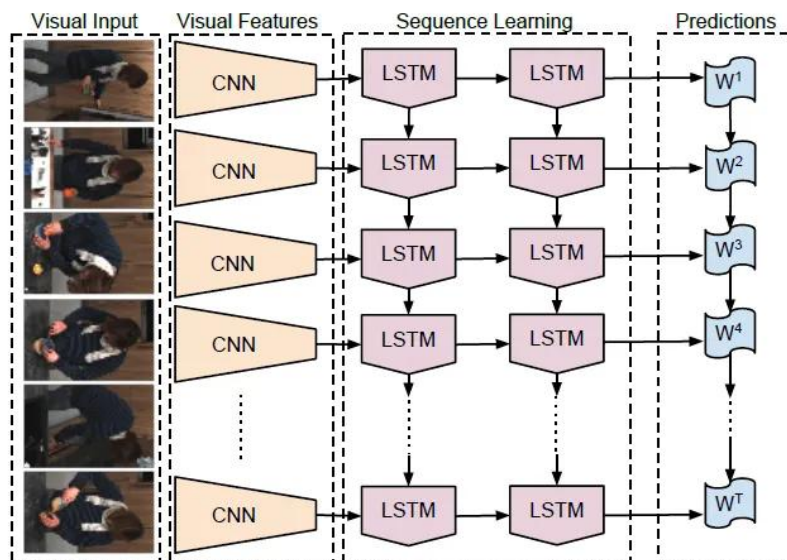


**Gambar 3.** ConvLSTM dalam penggunaan HAR

Dari gambar 3 tersebut, dapat dilihat bahwa metode ConvLSTM bekerja secara otomatis dalam mengenali aktivitas gerak manusia dengan memanfaatkan urutan gambar atau video frame untuk dikenali. Informasi terkait bagaimana tubuh bekerja dari waktu ke waktu direpresentasikan oleh setiap frame pada video dengan menggunakan posisi tubuh

manusia dalam waktu tertentu. Dengan menggunakan 42 filter dengan ukuran kernel  $3 \times 3$ , ConvLSTM dalam gambar memungkinkan model untuk mendeteksi pola gerakan lokal pada tubuh dari frame ke frame. Hasil dari proses ini adalah representasi fitur spasial-temporal yang kemudian diproses melalui lapisan yang penuh terhubung atau tebal. Tugas lapisan ini adalah menggabungkan data yang telah diekstraksi dan menyajikannya dalam bentuk yang dapat dimengerti oleh lapisan klasifikasi terakhir.

Selanjutnya, setelah selesai model dengan menggunakan metode ConvLSTM, langkah berikutnya adalah dengan melakukan pelatihan model dengan menggunakan metode LRCN. Berikut pada Gambar 4 akan ditampilkan bagaimana metode LRCN bekerja.



**Gambar 4.** LRCN dalam penggunaan HAR

Berdasarkan Gambar 4 diatas, Metode LRCN memulai proses pada rangkaian frame video yang dimasukkan sebagai input visual. Jaringan CNN memproses setiap frame untuk menghasilkan fitur spasial yang mewakili pose atau kondisi tubuh manusia setiap saat. Jaringan LSTM menggunakan hasil ekstraksi fitur dari seluruh frame untuk mempelajari hubungan temporal atau urutan gerakan antar frame. Dinamika aktivitas dapat ditangkap dalam rentang waktu tertentu. Pada akhirnya, pemahaman spasial-temporal yang diperoleh dari video tersebut digunakan untuk membuat prediksi aktivitas.

d. Evaluasi Performa Model








Setelah melakukan pelatihan terhadap kedua model yang akan digunakan, langkah selanjutnya adalah dengan melakukan evaluasi terhadap performa dari model yang telah dilatih. Evaluasi tersebut dilakukan dengan tujuan agar model dapat memprediksi atau mengenali video aktivitas gerak manusia dengan baik. Agar menghindari terjadinya




overfitting ataupun underfitting pada model yang telah dilatih, evaluasi dilakukan berdasarkan hasil dari train accuracy, validation accuracy, train loss, dan validation loss. Evaluasi tersebut akan ditampilkan dalam bentuk plot matriks berdasarkan masing-masing perbandingannya. Apabila model yang dilatih mengalami overfitting ataupun underfitting, maka model tersebut akan dilatih ulang dengan hyperparameter tuning hingga akurasi yang didapatkan menjadi lebih baik.

e. Download Video

Setelah melakukan pelatihan terhadap kedua model yang akan digunakan, langkah selanjutnya adalah dengan melakukan evaluasi terhadap performa dari model yang telah dilatih. Evaluasi tersebut dilakukan dengan tujuan agar model dapat memprediksi atau mengenali video aktivitas gerak manusia dengan baik. Agar menghindari terjadinya overfitting ataupun underfitting pada model yang telah dilatih, evaluasi dilakukan berdasarkan hasil dari train accuracy, validation accuracy, train loss, dan validation loss. Evaluasi tersebut akan ditampilkan dalam bentuk plot matriks berdasarkan masing-masing perbandingannya. Apabila model yang dilatih mengalami overfitting ataupun underfitting, maka model tersebut akan dilatih ulang dengan hyperparameter tuning hingga akurasi yang didapatkan menjadi lebih baik.

**Tabel 1.** Daftar Uji Coba Video

	Label	Gambar Video
1	PlayingGuitar	
2	PlayingGuitar	
3	YoYo	
4	YoYo	
5	Billiards	
6	Billiards	
7	Skateboarding	

8	Skateboarding	
9	Basketball	
10	Basketball	

Dari Tabel 1 tersebut dapat dilihat bahwa terdapat sepuluh total video yang akan dijadikan percobaan pengenalan pada kedua model yang telah dibuat. Video-video tersebut mencakup keseluruhan kelas yang telah ditentukan dengan masing-masing 2 video per kelas atau labelnya. Setelah video-video berhasil dikumpulkan, langkah selanjutnya yang peneliti lakukan adalah dengan melakukan prediksi pengenalan aksi terhadap video-video yang ada.

f. Prediksi Video

Selanjutnya langkah terakhir dalam penelitian pada perancangan sistem ini adalah dengan melakukan prediksi terhadap video-video yang telah diunduh dari youtube. Hasil dari prediksi video tersebut akan menampilkan kegiatan apa yang dilakukan pada video tersebut dan tingkat kepercayaan model terhadap prediksi yang dilakukan. Hasil prediksi video yang dilakukan akan dikatakan berhasil ketika tingkat kepercayaan yang dihasilkan melebihi angka 80% atau 0.80.

#### 4. HASIL DAN PEMBAHASAN

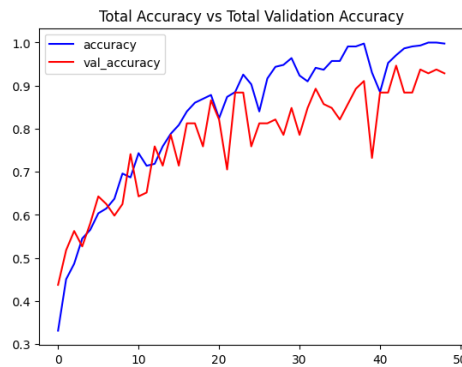
##### Hasil Evaluasi Performa Model

Pada bagian ini akan menjelaskan terkait bagaimana hasil evaluasi dari performa kedua model yang telah dibuat. Berikut adalah masing-masing hasil evaluasi performa dari kedua model tersebut.

a) Evaluasi Performa ConvLSTM Model

Pada tahapan ini akan dilakukan pembahasan terkait bagaimana hasil dari evaluasi pada performa model ConvLSTM yang telah dibuat. Model ini di lakukan training sebanyak 50 epoch dengan menggunakan EarlyStopping agar pelatihan otomatis berhenti ketika tidak ada peningkatan performa yang signifikan dari model. Dari total 50 epoch yang telah ditentukan, EarlyStopping menghentikan pelatihan dari model pada epoch ke-35 dengan hasil evaluasi akurasi model sebesar 0.8840 dan loss sebesar 0.3146.

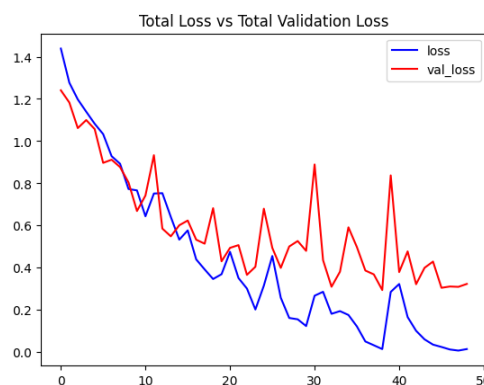
hasil ini menunjukkan bahwa model dilatih dengan cukup baik. Hasil tersebut juga dapat ditunjukkan berdasarkan akurasi dan loss dengan menggunakan plot matriks yang akan dilampirkan pada gambar dibawah ini.



**Gambar 5.** Plot Matriks Akurasi ConvLSTM Model

Dari gambar 5 tersebut dapat dilihat bahwa selama 35 epoch yang telah dilakukan, hasil pada training accuracy memiliki hasil yang sangat baik. Dari pelatihan ke-1 hingga ke-35 akurasi pelatihan pada model naik secara signifikan pada tiap epochnya dengan hanya terjadi beberapa kali penurunan yang tidak terlalu jauh. Hasil akhir yang dimiliki oleh model ConvLSTM pada epoch ke-35 sekitar 0.98 hasil training accuracy.

Berbeda dengan hasil pada training accuracy, hasil pada validation accuracy memiliki hasil yang kurang baik di setiap validasinya. Hasil tersebut menunjukkan bahwa akurasi dari validasi terjadi cukup fluktuatif, dimana di setiap pelatihannya akurasi yang di dapatkan tidak konsisten meningkat dan lebih sering terjadi akurasi yang naik dan turun. Fluktuatif inilah yang menyebabkan evaluasi dari performa model hanya memiliki akurasi sebesar 0.8840. Hasil akhir yang dimiliki oleh model ConvLSTM pada epoch ke-35 sekitar 0.90 hasil validation accuracy. Kemudian, berikut ini adalah penjelasan terkait hasil evaluasi pada plot matriks loss yang dilampirkan pada Gambar 6 berikut.



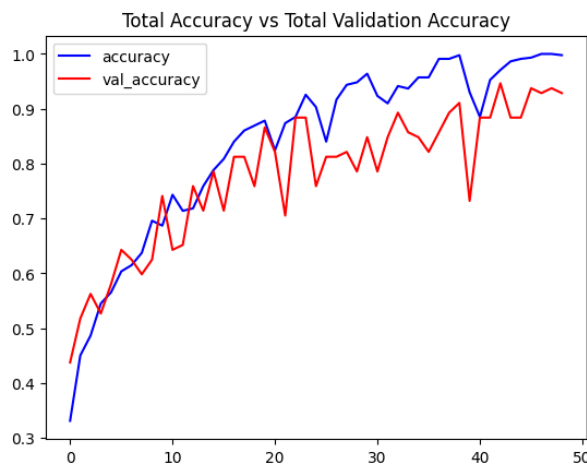
**Gambar 6.** Plot Matriks Loss ConvLSTM Model

Berdasarkan pada Gambar 6 diatas, hasil dari training loss dan validation loss sama-sama terjadi secara fluktuatif. Pada epoch awal, hasil pada training loss cukup baik yang ditandai dengan menurunnya loss secara signifikan. Namun, ketika jumlah epoch semakin meningkat, loss dari model terjadi penurunan dan kenaikan secara fluktuatif dengan hasil akhir dari loss tersebut sebesar 0,0536.

Kemudian pada bagian validation loss terjadi secara fluktuatif sejak awal awal epoch dilakukan. Hasil tersebut dapat ditunjukkan pada Gambar 6 tersebut dimana terjadi ketidakstabilan loss pada proses pelatihan model yang telah dibuat. Hasil akhir validation loss pada model ConvLSTM yang telah dibuat memiliki nilai sekitar 0,3349. dari hasil kedua plot matriks ini dapat ditentukan bahwa model telah dilatih dengan cukup baik dan dapat dilakukan uji coba prediksi pengenalan video.

b) Evaluasi Performa LRCN Model

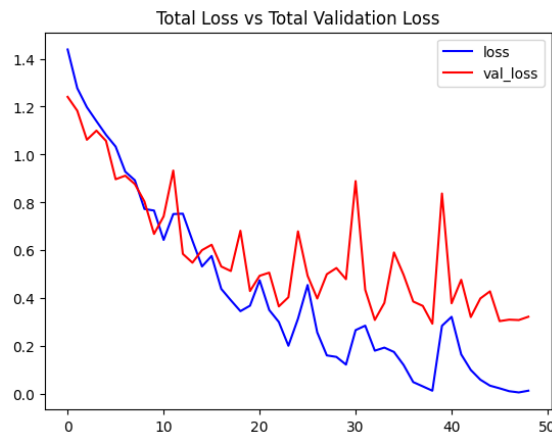
Selanjutnya, pada bagian ini akan membahas terkait bagaimana hasil evaluasi performa dari model LRCN. Sama seperti model ConvLSTM, Model ini juga di training sebanyak 50 epoch dengan menggunakan EarlyStopping agar pelatihan otomatis berhenti ketika tidak ada peningkatan performa yang signifikan dari model. Dari total 50 epoch yang telah ditentukan, EarlyStopping menghentikan pelatihan dari model pada epoch ke-49 dengan hasil evaluasi akurasi model sebesar 0,8837 dan loss sebesar 0,4216. jika dilihat dari hasil evaluasi pada model ini, hasil ini menunjukkan bahwa model LRCN yang telah dilatih memiliki akurasi dan loss yang sangat sedikit lebih buruk dari model ConvLSTM. Adapun daripada itu model LRCN ini juga dapat dikategorikan sebagai model yang memiliki performa yang cukup baik. Hasil tersebut juga dapat ditunjukkan berdasarkan akurasi dan loss dengan menggunakan plot matriks yang akan dilampirkan pada gambar dibawah ini.



Gambar 7. Plot Matriks Akurasi LRCN Model

Berdasarkan plot matriks pada Gambar 7 diatas, hasil pelatihan pada LRCN model menunjukkan akurasi yang sangat baik. Grafik yang ditunjukkan relatif naik sejak awal epoch dengan sedikit penurunan pada beberapa epoch nya. Hasil dari akurasi pelatihan sempat menyentuh akurasi sebesar 1,000 pada beberapa epoch namun diakhiri dengan nilai akurasi sebesar 0,9957 pada epoch ke-49.

Tidak sama dengan hasil training accuracy, grafik pada hasil validation accuracy menunjukkan bahwa model memiliki akurasi yang kurang baik. Dari grafik tersebut dapat dilihat bahwa terjadi fluktuasi secara signifikan dari awal hingga akhir epoch. Hal inilah yang menyebabkan model memiliki evaluasi dengan nilai yang hanya stagnan pada 0,8837. Hasil akhir epoch pada model LRCN ini memiliki nilai sekitar 0,92. Adapun berikut ini adalah penjelasan terkait hasil evaluasi pada plot matriks loss yang dilampirkan pada Gambar 8 berikut.



**Gambar 8.** Plot Matriks Loss LRCN Model

Sama seperti grafik yang terjadi pada plot matriks akurasi, grafik yang ada pada Gambar 8 diatas menunjukkan bahwa hasil training loss yang ada pada model LRCN mengalami fluktuasi seiring dengan meningkatnya epoch. Pada awal epoch, training loss menurun secara signifikan dan mulai mengalami fluktuasi pada pertengahan hingga akhir epoch. Hasil akhir dari training loss tersebut memiliki nilai sekitar 0,17.

Sedangkan pada bagian validation loss, fluktuasi terjadi dari awal hingga akhir epoch. Pada beberapa epoch terjadi kenaikan yang cukup tinggi yang menunjukkan bahwa model tidak cukup baik dalam melakukan pembelajaran terhadap data yang ada. Hasil akhir dari validation loss memiliki nilai sekitar 0.32. Hal inilah yang menyebabkan tingginya nilai dari loss pada evaluasi model yang telah dilakukan.

c) Prediksi Pengenalan Aktivitas

Pada tahapan ini akan dilakukan pembahasan pada prediksi video yang telah dilakukan. Prediksi tersebut dilakukan pada 10 sampel video yang di download dari TikTok dengan masing masing 2 video pada setiap kelasnya. Berikut ini adalah hasil dari prediksi pada keseluruhan video dengan menggunakan kedua model yang telah dilatih yang akan dilampirkan pada Tabel 2 berikut ini.

**Tabel 2.** Hasil Prediksi

Label	Hasil Prediksi		Confidence	
	ConvLSTM	LRCN	ConvLSTM	LRCN
PlayingGuitar	SkateBoarding	PlayingGuitar	0.96	0.92
PlayingGuitar	PlayingGuitar	PlayingGuitar	0.96	0.97
YoYo	SkateBoarding	YoYo	0.79	0.84
YoYo	PlayingGuitar	YoYo	0.97	0.99
Billiards	SkateBoarding	Billiards	0.96	0.99
Billiards	SkateBoarding	Billiards	0.95	0.99
SkateBoarding	Basketball	SkateBoarding	0.34	0.99
SkateBoarding	Basketball	SkateBoarding	0.50	0.99
Basketball	Basketball	Basketball	0.97	0.93
Basketball	SkateBoarding	Basketball	0.64	0.87

Berdasarkan data hasil prediksi yang ada pada Tabel 2 diatas, hasil menunjukkan bahwa model LRCN memiliki performa yang sangat baik dalam mengenali aktivitas yang ada pada video hasil unduhan dari platform TikTok tersebut. Dari total 10 video yang telah disediakan, LRCN model berhasil memprediksi dan mengenali semua aktivitas yang ada di video tersebut berdasarkan kelas-kelasnya dengan tingkat confidence konsisten diatas 80%.

Berbeda dengan model LRCN, ConvLSTM model hanya berhasil memprediksi aktivitas pada video sebanyak 2 prediksi benar dari total 10 video yang ada. ConvLSTM Model cenderung memprediksi pengenalan aktivitas gerak manusia dengan salah dengan tingkat confidence yang tinggi. ConvLSTM model beberapa kali memprediksi video dengan salah dengan tingkat confidence diatas 90%. Dengan adanya hasil prediksi ini, meskipun ConvLSTM model memiliki hasil evaluasi performa yang lebih baik, namun LRCN model memiliki performa yang jauh lebih baik dalam memprediksi pengenalan aktivitas gerak manusia. LRCN model akan jauh lebih cocok ketika digunakan untuk mengenali aktivitas gerak manusia dengan pelatihan model yang perlu ditingkatkan lagi.

#### 4. KESIMPULAN DAN SARAN

Dalam artikel yang membandingkan kemampuan model ConvLSTM dan LRCN dalam pengenalan aktivitas manusia dari data video, dapat disimpulkan bahwa kedua model menunjukkan kemampuan yang cukup baik dalam proses pelatihan, dengan ConvLSTM mencapai akurasi pelatihan sekitar 98% dan LRCN mencapai akurasi pelatihan hingga 99,57%. Namun, dalam pengujian nyata menggunakan video dari platform seperti TikTok, LRCN menunjukkan keunggulan yang signifikan dalam pengenalan semua aktivitas manusia. Oleh karena itu, untuk pengenalan aktivitas manusia dari video, model LRCN lebih baik, meskipun tuning hyperparameter serta augmentasi data diperlukan untuk meningkatkan kinerja.

Saran untuk pengembangan selanjutnya adalah dengan melakukan tuning hyperparameter dan penerapan teknik augmentasi data disarankan untuk membuat model pengenalan aktivitas manusia lebih kuat dan akurat. Untuk memastikan bahwa model dapat diterapkan dalam berbagai kondisi, pengujian pada berbagai dataset dan kondisi nyata sangat penting. Pengembangan arsitektur yang lebih kompleks atau transfer learning juga dapat membantu sistem pengenalan aktivitas menjadi lebih andal.

#### DAFTAR REFERENSI

- Ahmad, I., Ullah, F., Khan, M. A., & Kim, D. (2021). Human action recognition in smart surveillance using hybrid deep learning model. *Sensors*, 21(18), 6104. <https://doi.org/10.3390/s21186104>
- Almars, A. M. (2021). Deepfakes detection techniques using deep learning: A survey. *Journal of Computer and Communications*, 9(5), 20–35. <https://doi.org/10.4236/jcc.2021.95003>
- Arshad, M. H., Bilal, M., & Gani, A. (2022). Human activity recognition: Review, taxonomy and open challenges. *Sensors*, 22(17), 1–33. <https://doi.org/10.3390/s22176463>
- Elmetwally, A., Eldeeb, R., & Elmougy, S. (2025). Deep learning based anomaly detection in real-time video. *Multimedia Tools and Applications*, 84(11), 9555–9571. <https://doi.org/10.1007/s11042-024-19116-9>
- Gupta, S. (2021). Deep learning based human activity recognition (HAR) using wearable sensor data. *International Journal of Information Management Data Insights*, 1(2), 100046. <https://doi.org/10.1016/j.ijime.2021.100046>
- Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., & Davis, L. S. (2016). Learning temporal regularity in video sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 733–742). <https://doi.org/10.1109/CVPR.2016.85>
- Ismail, A. P., Aziz, F. A. A., Kasim, N. M., & Daud, K. (2021). Hand gesture recognition on Python and OpenCV. *IOP Conference Series: Materials Science and Engineering*, 1045(1), 012043. <https://doi.org/10.1088/1757-899x/1045/1/012043>

- Komatsu, M., Sakai, A., Komatsu, R., Matsuoka, R., Yasutomi, S., Shozu, K., ... Hamamoto, R. (2021). Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. *Applied Sciences (Switzerland)*, *11*(1), 1–12. <https://doi.org/10.3390/app11010371>
- Lina, Augustine, M., Stephen, R., & Salim, L. (2024). Pengenalan aktivitas manusia dalam ruangan dengan convolutional neural networks. *Teknika*, *13*(1), 58–64. <https://doi.org/10.34148/teknika.v13i1.707>
- Mansour, R. F., Escorcía-Gutiérrez, J., Gamarra, M., Villanueva, J. A., & Leal, N. (2021). Intelligent video anomaly detection and classification using Faster R-CNN with deep reinforcement learning model. *Image and Vision Computing*, *112*, 104229. <https://doi.org/10.1016/j.imavis.2021.104229>
- Rahman, A., Islam, M., Moon, M. J., Tasnim, T., & Siddique, N. (2022). A qualitative survey on deep learning-based deepfake video creation and detection method. *Australian Journal of Engineering and Innovative Technology*, 13–26. <https://doi.org/10.34104/ajeit.022.013026>
- Ravanbakhsh, M., Nabi, M., Mousavi, H. S., & Sebe, N. (2017). Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 1577–1581). <https://doi.org/10.1109/ICIP.2017.8296547>
- Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6479–6488). <https://doi.org/10.1109/CVPR.2018.00678>
- Uddin, M. A., Talukder, M. A., Uzzaman, M. S., Debnath, C., Chanda, M., Paul, S., ... Aryal, S. (2024). Deep learning-based human activity recognition using CNN, ConvLSTM, and LRCN. *International Journal of Cognitive Computing in Engineering*, *5*, 259–268. <https://doi.org/10.1016/j.ijcce.2024.06.004>
- Vrskova, R., Hudec, R., Kamencay, P., & Sykora, P. (2022). A new approach for abnormal human activities recognition based on ConvLSTM architecture. *Sensors*, *22*(8), 1–20. <https://doi.org/10.3390/s22082946>
- Wu, P., Pan, C., Yan, Y., Pang, G., Wang, P., & Zhang, Y. (2024). Deep learning for video anomaly detection: A review. *arXiv*. <http://arxiv.org/abs/2409.05383>
- Zhou, T., Porikli, F., Crandall, D. J., Van Gool, L., & Wang, W. (2023). A survey on deep learning technique for video segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(6), 7099–7122. <https://doi.org/10.1109/TPAMI.2022.3225573>