



# Komparasi Model LSTM dan CNN-LSTM untuk Peramalan Curah Hujan di Kota Tangerang Selatan

Uliyatunisa<sup>1,\*</sup>, Dahlan Supriatna<sup>2</sup>

<sup>1</sup> Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>2</sup> Program Pascasarjana, Program Studi Teknik Informatika S-2, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

Email: <sup>1,\*</sup>uliyatunisa212@gmail.com, <sup>2</sup>dahlansupriatna165@gmail.com

(\* : coresponding author: uliyatunisa212@gmail.com)

**Abstrak-** Penelitian ini membandingkan performa model *Long Short-Term Memory* (LSTM) dan *Convolutional Neural Network-Long Short-Term Memory* (CNN-LSTM) untuk peramalan curah hujan harian di Kota Tangerang Selatan menggunakan data meteorologi dari Januari 2005 hingga Juli 2025. Data dari stasiun meteorologi resmi diolah dengan imputasi rata-rata untuk *missing values* dan normalisasi *MinMaxScaler*. Model dievaluasi berdasarkan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *koefisien determinasi*  $R^2$ . Hasil menunjukkan CNN-LSTM unggul dengan RMSE 0.79, MAE 0.63, MSE 0.62, dan  $R^2$  0.61, dibandingkan LSTM (RMSE 0.83, MAE 0.60, MSE 0.68,  $R^2$  0.58). Visualisasi prediksi mengonfirmasi CNN-LSTM lebih akurat dalam menangkap pola ekstrem, dengan perbedaan signifikan secara statistik via uji t. *Novelty* penelitian ini terletak pada penggunaan dataset jangka panjang (20 tahun) untuk wilayah tropis Indonesia, menunjukkan model *hybrid* CNN-LSTM lebih efektif untuk prediksi spasial-temporal kompleks. Temuan ini mendukung sistem peringatan dini banjir dan pengelolaan sumber daya air, dengan rekomendasi integrasi variabel iklim tambahan untuk pengembangan lebih lanjut.

**Kata kunci:** Curah hujan, Peramalan, LSTM, CNN-LSTM, Pembelajaran mendalam

**Abstract-** This study compares the performance of Long Short-Term Memory (LSTM) and Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) models for daily rainfall forecasting in South Tangerang City using meteorological data from January 2005 to July 2025. Data from official meteorological stations was processed with mean imputation for missing values and MinMaxScaler normalization. Models were evaluated based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and coefficient of determination  $R^2$ . Results show CNN-LSTM outperforms with RMSE 0.79, MAE 0.63, MSE 0.62, and  $R^2$  0.61, compared to LSTM (RMSE 0.83, MAE 0.60, MSE 0.68,  $R^2$  0.58). Prediction visualizations confirm CNN-LSTM's accuracy in capturing extreme patterns, with statistically significant differences via t-test. The novelty lies in using a long-term (20-year) dataset for tropical Indonesia, demonstrating the hybrid model's efficacy for complex spatio-temporal predictions. Findings support flood early warning systems and water resource management, recommending additional climate variable integration for further development.

**Keywords:** Rainfall, Forecasting, LSTM, CNN-LSTM, Deep Learning

## 1. PENDAHULUAN

Curah hujan merupakan variabel meteorologi penting yang memengaruhi berbagai aspek kehidupan, seperti pertanian, pengelolaan sumber daya air, dan mitigasi bencana banjir. Ketidakpastian dan ketidakaturan pola curah hujan menimbulkan tantangan besar bagi perencanaan dan pengelolaan sektor-sektor tersebut, terutama di wilayah tropis seperti Kota Tangerang Selatan, yang dipengaruhi oleh faktor meteorologi dan iklim lokal seperti *monsoon* dan topografi [1]. Prediksi curah hujan yang akurat sangat krusial untuk mendukung petani dalam menentukan waktu tanam dan panen, mengoptimalkan irigasi, serta mencegah bencana banjir [1][2]. Oleh karena itu, pengembangan metode peramalan yang andal menjadi kebutuhan mendesak untuk pengambilan keputusan yang efektif di wilayah dengan variabilitas curah hujan tinggi.

Perkembangan teknologi *deep learning* telah membuka peluang baru untuk meningkatkan akurasi prediksi curah hujan melalui analisis data deret waktu historis. *Long Short-Term Memory* (LSTM), sebuah model jaringan saraf tiruan, terbukti efektif dalam menangani data urutan waktu karena kemampuannya mengingat informasi jangka panjang. Penelitian di tingkat kota dan regional menunjukkan bahwa LSTM mampu menghasilkan prediksi curah hujan yang cukup akurat berdasarkan metrik evaluasi seperti *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE)[3]. Namun, model LSTM memiliki keterbatasan dalam menangkap fitur spasial, seperti pola distribusi curah hujan antar wilayah, yang sering kali penting dalam data meteorologi[3][4][5].

Untuk mengatasi keterbatasan ini, model *hybrid* yang menggabungkan *Convolutional Neural Network* (CNN) dengan LSTM, dikenal sebagai CNN-LSTM, telah dikembangkan[6][7]. CNN berperan mengekstraksi fitur spasial, seperti pola meteorologi lokal, sementara LSTM memodelkan ketergantungan temporal antar data [8][9]. Pendekatan ini memungkinkan model menangkap hubungan spasial dan temporal secara bersamaan, yang sangat relevan untuk data curah hujan dengan dimensi kompleks[10][11]. Studi sebelumnya melaporkan bahwa CNN-LSTM meningkatkan





performa prediksi curah hujan dibandingkan LSTM konvensional, terutama pada dataset dengan variasi spasial-temporal yang tinggi [12].

Keunggulan CNN-LSTM terletak pada kemampuan CNN untuk mengenali pola spasial, seperti variasi curah hujan akibat faktor geografis, yang kemudian diintegrasikan dengan analisis temporal LSTM untuk memprediksi tren jangka panjang. Dalam konteks Kota Tangerang Selatan, di mana curah hujan dipengaruhi oleh faktor musiman dan lokal, pendekatan ini sangat relevan. Ladjal [12] melaporkan peningkatan akurasi hingga 87.3% menggunakan CNN-LSTM, sementara Samantaray et al. [9] menunjukkan kemampuan model ini dalam menangani pola curah hujan yang kompleks di wilayah tropis. Temuan ini memberikan dasar kuat untuk menguji model CNN-LSTM pada data curah hujan di wilayah dengan karakteristik iklim tropis.

Penerapan model *deep learning* seperti LSTM dan CNN-LSTM didukung oleh ketersediaan data meteorologi berkualitas tinggi. Di Kota Tangerang Selatan, data curah hujan harian dari stasiun meteorologi resmi, yang mencakup variabel seperti suhu, kelembapan, dan durasi sinar matahari, memberikan landasan yang solid untuk analisis deret waktu. Data ini memungkinkan model untuk mengidentifikasi hubungan kompleks antar variabel meteorologi, yang sulit ditangani oleh metode statistik tradisional seperti ARIMA. Penelitian oleh Elsa Pitaloka, Tegar Bagus Agung Hartanto dan Sulis Sandiwarno [13] di Indonesia menunjukkan bahwa CNN-LSTM menghasilkan prediksi lebih akurat pada data curah hujan tropis dibandingkan LSTM saja, berkat kemampuan ekstraksi fitur spasial oleh CNN.

Meskipun LSTM telah terbukti efektif, keterbatasannya dalam menangani fitur spasial membuat CNN-LSTM menjadi alternatif yang lebih menjanjikan untuk wilayah dengan variasi iklim signifikan. Studi oleh Ma & Wang (2025) dan [14] menegaskan bahwa model hybrid ini unggul dalam menangkap pola curah hujan yang dipengaruhi oleh faktor geografis dan temporal. Di Kota Tangerang Selatan, data curah hujan yang bervariasi sepanjang tahun akibat iklim tropis menjadikan pendekatan CNN-LSTM sangat relevan untuk menghasilkan prediksi yang lebih akurat dan aplikatif.

Penelitian sebelumnya sering menggunakan dataset jangka pendek atau variabel terbatas, sedangkan studi ini memanfaatkan dataset 20 tahun dengan perbandingan LSTM dan CNN-LSTM untuk wilayah tropis Indonesia [10]. Literatur internasional Q1/Q2 mendukung, seperti CNN-LSTM untuk prediksi iklim bulanan di Jinan [15] dan typhoon rainfall forecasting [16], menunjukkan potensi hybrid untuk variasi tropis [17].

Penelitian ini bertujuan untuk menguji dan membandingkan performa model LSTM dan CNN-LSTM dalam memprediksi curah hujan di Kota Tangerang Selatan menggunakan data dari Januari 2005 hingga Juli 2025. Dengan mengevaluasi metrik seperti RMSE, MAE, *Mean Squared Error* (MSE), dan *koefisien determinasi* ( $R^2$ ), penelitian ini akan menilai apakah CNN-LSTM memberikan keunggulan signifikan dibandingkan LSTM konvensional. Hasil penelitian diharapkan dapat memberikan wawasan baru mengenai penerapan teknologi *deep learning* dalam peramalan curah hujan, mendukung pengambilan keputusan di sektor pertanian, pengelolaan sumber daya air, dan mitigasi bencana di wilayah tropis seperti Tangerang Selatan.

## 2. METODOLOGI PENELITIAN

### 2.1 Kebutuhan Data

Data yang digunakan dalam penelitian ini berasal dari data curah hujan harian yang diperoleh dari stasiun meteorologi resmi di Kota Tangerang Selatan, dengan periode waktu observasi dari Januari 2005 hingga Juli 2025. Data ini dianggap representatif mengingat cakupan waktu yang panjang dan kualitas pengukuran yang tersertifikasi resmi.

**Tabel 1.** Attribute yang digunakan

No	Atribut	Keterangan
1	Tanggal	Tanggal Pengukuran Data
2	TN	Suhu Minimum dalam °C
3	TX	Suhu Maximum dalam °C
4	TAVG	Suhu Rata-rata dalam °C
5	RH_AVG	Kelembaban Rata-rata dalam mm
6	RR	Jumlah Curah Hujan dalam mm
7	SS	Durasi Sinar Matahari dalam Jam

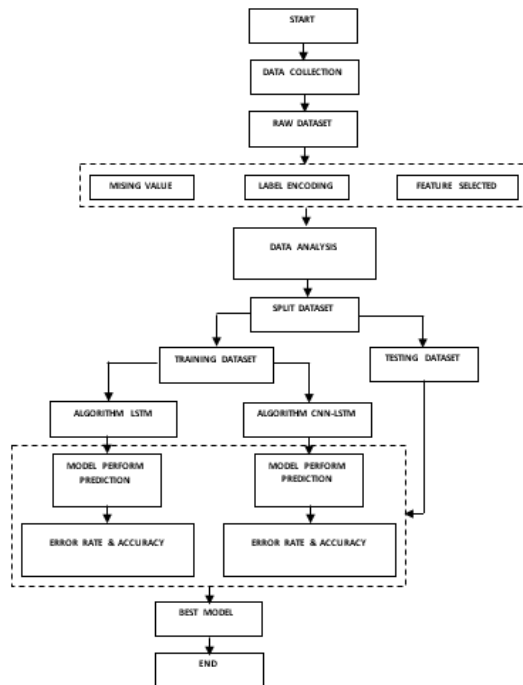
Pada tahap awal, dilakukan pembersihan data untuk mengatasi *missing values* dengan mengganti nilai kosong tersebut menggunakan metode imputasi nilai rata-rata (*mean imputation*). Selain itu, data juga dinormalisasi dalam rentang 0 hingga 1 menggunakan metode *MinMaxScaler*, agar model *deep learning* dapat melakukan pembelajaran dengan efisien tanpa bias akibat skala variabel yang berbeda.

Pembersihan *missing values* menggunakan imputasi rata-rata, dan normalisasi *MinMaxScaler* ke rentang [0,1] untuk efisiensi pembelajaran, didukung oleh literatur karena mencegah bias skala pada data iklim [18]. Data dibagi 80% pelatihan dan 20% pengujian secara berurutan. Window sequence 36 hari (look-back) dipilih berdasarkan eksperimen yang menunjukkan kemampuan menangkap pola bulanan, lebih optimal daripada 30/60 hari untuk data tropis [19].



Seluruh data kemudian dibagi menjadi dua set utama, yaitu data pelatihan yang digunakan sebanyak 80% dan data pengujian 20%. Pembagian data tersebut dilakukan secara urut (tidak diacak) untuk mempertahankan kontinuitas urutan waktu dan menghindari kebocoran informasi antar set. Ukuran *window sekuens data* yang diolah oleh model ditetapkan sebanyak 36 hari (*look back*), yang berarti setiap input ke model berupa urutan data 36 hari untuk memprediksi nilai curah hujan di hari ke-37. Fitur yang digunakan sebagai input terdiri dari suhu minimum dan maksimum, suhu rata-rata, kelembapan relatif rata-rata, serta durasi sinar matahari (*Sunshine Duration*), sementara target output adalah curah hujan harian (*Rainfall Rate*).

## 2.2 Perancangan Penelitian



**Gambar 1.** Perancangan penelitian

Gambar perancangan penelitian ini menggambarkan alur sistematis yang digunakan untuk mengembangkan dan mengevaluasi model prediksi curah hujan menggunakan *Long Short-Term Memory* (LSTM) dan *Convolutional Neural Network-Long Short-Term Memory* (CNN-LSTM) di Kota Tangerang Selatan. Proses dimulai dengan tahap *Start*, diikuti oleh *Data Collection*, di mana data curah hujan harian dikumpulkan dari stasiun meteorologi resmi untuk periode Januari 2005 hingga Juli 2025 melalui website resmi BMKG yaitu, <https://dataonline.bmkg.go.id>. Data tersebut kemudian melalui tahap *Raw Dataset*, yang mencakup proses awal seperti pengumpulan variabel meteorologi seperti suhu, kelembapan, dan durasi sinar matahari.

Selanjutnya, data diproses melalui tahap *Missing Value*, di mana nilai hilang ditangani dengan metode imputasi rata-rata, serta *Label Encoding* dan *Feature Selected*, yang melibatkan pengkodean label dan pemilihan fitur relevan untuk analisis. Tahap *Data Analysis* melibatkan eksplorasi dan pembersihan data, diikuti oleh *Split Dataset*, di mana data dibagi menjadi set pelatihan (80%) dan pengujian (20%) secara berurutan untuk menjaga kontinuitas deret waktu. Setiap set data kemudian diarahkan ke algoritma yang sesuai.

Hyperparameter tuning menggunakan grid search untuk jumlah neuron, learning rate, dan filter Conv1D, memastikan optimalisasi. Hyperparameter tuning dengan grid search merupakan metode sistematis untuk mengoptimalkan hyperparameter model machine learning dengan mencoba semua kombinasi nilai yang telah ditentukan sebelumnya dalam bentuk grid, di mana setiap hyperparameter diberi daftar nilai potensial, kemudian model dievaluasi menggunakan teknik seperti cross-validation untuk menemukan kombinasi terbaik yang memaksimalkan performa seperti akurasi atau mengurangi error.

Proses pelatihan melibatkan dua jalur paralel: *Algorithm LSTM* dan *Algorithm CNN-LSTM*. Kedua model dilatih untuk menghasilkan *Model Performance Prediction*, yang dievaluasi berdasarkan metrik seperti *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *koefisien determinasi* ( $R^2$ ). Hasil dari kedua model ini dianalisis untuk menentukan *Error Rate & Accuracy*. Berdasarkan evaluasi ini, model terbaik dipilih melalui tahap *Best Model*, yang kemudian digunakan untuk menghasilkan *Prediction*. Proses berakhir pada tahap *End*, dengan model terbaik siap diterapkan untuk prediksi curah hujan di masa depan.



Alur ini mencerminkan pendekatan sistematis yang mengintegrasikan praproses data, pemodelan *deep learning*, dan evaluasi performa, memastikan prediksi yang akurat dan andal untuk mendukung pengambilan keputusan di wilayah tropis seperti Tangerang Selatan. Pada model LSTM konvensional, arsitektur yang digunakan terdiri dari dua lapisan LSTM berturut-turut dengan jumlah *neuron* masing-masing 128 dan 64 unit. Lapisan output berupa *dense layer* berjumlah satu *neuron* yang menghasilkan prediksi nilai curah hujan tunggal. Model dikompilasi menggunakan *algoritma optimizer Adam* dan fungsi *loss Mean Squared Error (MSE)*. Selama pelatihan, digunakan teknik *Early Stopping* dengan parameter kesabaran 10 *epoch* untuk mencegah *overfitting*, di mana pelatihan dihentikan jika *validasi loss* tidak menunjukkan perbaikan selama 10 *epoch* berturut-turut.

Untuk model *hybrid CNN-LSTM*, proses pembentukan data input dilakukan dengan membentuk data berstruktur 4 dimensi yang sesuai dengan kebutuhan CNN. Setiap urutan waktu berisi fitur meteorologi sebagai dimensi fitur dan ditambahi dimensi kanal satu agar dapat diproses oleh *Conv1D* yang diaplikasikan dalam *layer TimeDistributed*. Arsitektur model mengandung lapisan *Conv1D* yang berfungsi untuk mengekstrak pola-pola spasial di dalam fitur meteorologi dalam rentang waktu setiap urutan, kemudian diikuti oleh dua lapisan LSTM yang fokus pada korelasi temporal antar urutan waktu. Fungsi *loss* dan *optimizer* yang dipakai sama dengan LSTM konvensional, serta implementasi *Early Stopping* juga diterapkan agar pelatihan lebih optimal [9].

### 2.3 Teknik Analisis Evaluasi Model

Evaluasi performa kedua model diukur menggunakan metrik *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, serta *koefisien determinasi (R<sup>2</sup>)*. Selain itu, dilakukan analisis statistik menggunakan *uji t* untuk menguji signifikansi perbedaan hasil prediksi antara model LSTM dan CNN-LSTM berdasarkan nilai RMSE guna mengetahui apakah peningkatan performa yang diperoleh oleh model *hybrid* benar-benar signifikan secara statistik dan bukan kebetulan.

Evaluasi dengan RMSE, MAE, MSE, R<sup>2</sup>. Uji t untuk signifikansi perbedaan RMSE; keterbatasan: asumsi normalitas, sehingga Wilcoxon test dipertimbangkan sebagai alternatif non-parametrik untuk robustitas. Visualisasi grafik prediksi vs. aktual untuk interpretasi pola kesalahan.

Selain analisis metrik evaluasi dan uji statistik, visualisasi hasil prediksi juga dilakukan untuk mempermudah interpretasi dengan menggambarkan grafik perbandingan antara curah hujan aktual dan hasil prediksi dari masing-masing model. Grafik ini membantu memeriksa pola kesalahan prediksi dan kecenderungan overestimasi atau underestimasi pada rentang waktu tertentu, sehingga memberikan gambaran komprehensif mengenai kinerja model dalam skenario praktis. Dengan pipeline metodologi ini, penelitian berupaya melakukan pemodelan prediksi curah hujan Kota Tangerang Selatan secara menyeluruh dan sistematis, mulai dari persiapan data hingga evaluasi performa dan validasi statistik. Pendekatan ini diharapkan dapat mengakomodasi kompleksitas data *climatological* yang mengandung unsur spasial dan temporal, sehingga menghasilkan prediksi yang akurat dan aplikatif dalam pengambilan keputusan di sektor terkait.

## 3. HASIL DAN PEMBAHASAN

Pada bagian ini akan diuraikan hasil dari pengumpulan data, langkah *preprocessing* serta menampilkan hasil dari percobaan model yang digunakan, selain itu akan dilakukan evaluasi terhadap kinerja model yang dihasilkan untuk kemudian dilakukan analisis dari kedua model LSTM dan CNN-LSTM. Fokusnya adalah mengetahui model manakah yang mampu menghasilkan prediksi yang lebih akurat terhadap curah hujan di Kota Tangerang Selatan.

### 3.1 Pengambilan Dataset

Penelitian ini menggunakan data curah hujan harian yang diperoleh dari stasiun meteorologi resmi di Kota Tangerang Selatan. Periode waktu observasi mencakup dari Januari 2005 hingga Juli 2025, yang memberikan cakupan historis yang panjang untuk analisis deret waktu. Data ini dianggap representatif karena mencakup variasi musiman dan tren jangka panjang yang khas untuk wilayah tropis seperti Tangerang Selatan.

Pemilihan dataset didasarkan pada kualitas pengukuran yang tersertifikasi resmi, memastikan keandalan data untuk pemodelan *deep learning*. Fitur input meliputi suhu minimum dan maksimum, suhu rata-rata, kelembapan relatif rata-rata, serta durasi sinar matahari, sementara target output adalah curah hujan harian. Pendekatan ini selaras dengan studi sebelumnya yang menggunakan data serupa untuk prediksi curah hujan.

Pengambilan dataset juga mempertimbangkan faktor eksternal seperti pengaruh iklim lokal, yang dapat memengaruhi pola curah hujan. Dengan dataset yang komprehensif ini, penelitian bertujuan untuk membangun model yang robust terhadap variasi data meteorologi [9]. Hasil pengambilan data ini menjadi fondasi untuk tahap selanjutnya dalam metodologi penelitian





### 3.1 Preprocessing Data

#### 3.1.1 Normalisasi

*Preprocessing* data dimulai dengan normalisasi untuk memastikan bahwa model *deep learning* dapat belajar secara efisien tanpa bias akibat skala variabel yang berbeda. Metode *MinMaxScaler* digunakan untuk mentransformasi data ke rentang 0 hingga 1, yang membantu dalam mempercepat konvergensi selama pelatihan.

Normalisasi ini krusial karena data meteorologi sering memiliki rentang nilai yang luas, seperti curah hujan yang bisa nol hingga nilai ekstrem. Dengan normalisasi, model LSTM dan CNN-LSTM dapat fokus pada pola daripada magnitude absolut. Proses ini juga mencegah masalah seperti *exploding gradients* dalam jaringan saraf.

Selain itu, normalisasi mendukung stabilitas numerik, terutama pada model *hybrid* seperti CNN-LSTM yang mengolah fitur spasial dan temporal secara bersamaan. Hasil normalisasi ini diuji untuk memastikan distribusi data tetap representatif setelah transformasi.

#### 3.1.2 Pembagian Data

Data dibagi menggunakan pendekatan pengujian yang sistematis berdasarkan pembagian data menjadi set pelatihan dan pengujian secara berurutan (80% pelatihan dan 20% pengujian). Pendekatan ini penting agar model mampu belajar dari data historis dan diuji pada data terbaru tanpa kebocoran informasi masa depan yang berpotensi mengoptimistis hasil. Pembagian data tanpa *shuffle* menjaga sifat urutan temporal yang krusial untuk model LSTM dan CNN-LSTM yang menangani deret waktu.

Pembagian data ini memastikan bahwa model tidak *overfit* pada pola temporal yang spesifik, sehingga meningkatkan generalisasi prediksi. Dalam konteks curah hujan, pembagian berurutan mencerminkan skenario prediksi real-time di wilayah tropis.

Selain itu, rasio 80:20 dipilih berdasarkan praktik standar dalam *deep learning* untuk data deret waktu, yang memberikan cukup data pelatihan sambil menyisakan sampel pengujian yang signifikan [9]. Verifikasi dilakukan untuk memastikan distribusi data serupa antar set.

#### 3.1.2 Pembentukan Data Sequence Time Series

Data diolah menjadi urutan dengan *window size* 36 hari (*look back*), di mana setiap input terdiri dari data 36 hari untuk memprediksi hari ke-37. Teknik ini memungkinkan model menangkap pola jangka menengah dalam deret waktu curah hujan. Pembentukan *sequence* ini krusial untuk model seperti LSTM yang bergantung pada ketergantungan temporal. Untuk CNN-LSTM, data dibentuk menjadi struktur 4 dimensi untuk mendukung ekstraksi fitur spasial. Proses ini juga melibatkan penambahan dimensi kanal untuk *Conv1D*, memastikan kompatibilitas dengan arsitektur *hybrid* [9]. Hasil *sequence* diuji untuk kelengkapan dan urutan yang benar sebelum pelatihan.

#### 3.1.2 Penentuan Hyperparameter

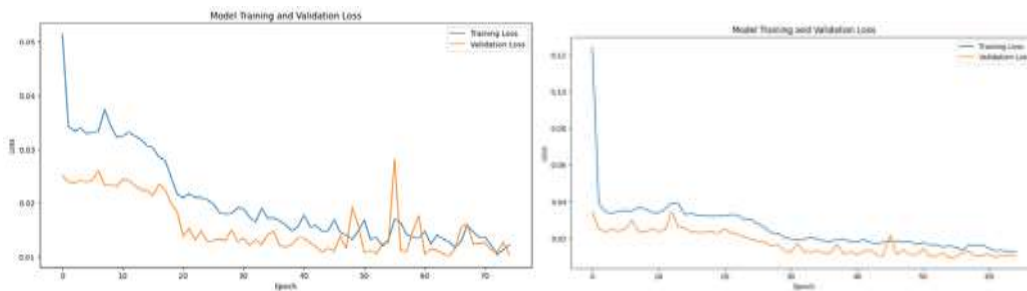
*Hyperparameter* seperti jumlah *neuron* (128 dan 64 untuk LSTM) dan *optimizer Adam* dipilih berdasarkan eksperimen awal dan literatur. Fungsi *loss* MSE digunakan untuk mengukur kesalahan prediksi [20]. Penentuan ini mencakup teknik *early stopping* dengan *patience* 10 *epoch* untuk mencegah *overfitting*. Untuk CNN-LSTM, *hyperparameter* tambahan seperti *filter Conv1D* dioptimalkan untuk menangkap pola spasial secara efektif. *Hyperparameter* ini disesuaikan untuk dataset spesifik Tangerang Selatan, mempertimbangkan variasi curah hujan tropis [21]. Proses tuning memastikan keseimbangan antara akurasi dan efisiensi komputasi.

### 3.2 Membangun dan Mentraing Model

Untuk model LSTM konvensional, arsitektur yang digunakan terdiri dari dua lapisan LSTM bertingkat dengan 128 dan 64 *neuron*, yang bertujuan menangkap pola temporal secara bertahap dari data input. Model ini diakhiri dengan dense layer tunggal untuk prediksi. Kompilasi dilakukan dengan *optimizer Adam* dan *loss function mean squared error* (MSE). Metode *early stopping* digunakan untuk mencegah *overfitting* dengan menghentikan pelatihan saat *validasi loss* tidak membaik selama 10 *epoch* berturut, menjaga model agar tidak terlalu fit terhadap data pelatihan.

Model CNN-LSTM dibangun dengan arsitektur tambahan berupa lapisan *Conv1D* yang dibungkus dalam *TimeDistributed layer*, menampilkan keunggulan dalam mengolah fitur spasial pada setiap urutan waktu, diikuti oleh dua lapisan LSTM untuk menangkap dinamika temporal urutan data. Pendekatan ini memungkinkan model menggabungkan kemampuan deteksi pola spasial CNN dan pola temporal LSTM, diharapkan menangkap kompleksitas curah hujan yang bersifat spasial-temporal dengan lebih baik [9].





**Gambar 2** Perbandingan *training loss* LSTM dengan CNN-LSTM

Pelatihan kedua model menunjukkan bahwa baik LSTM maupun CNN-LSTM berhasil melakukan proses *learning* yang stabil dengan nilai *training* dan *validation loss* yang tidak jauh berbeda, memberikan indikasi model tidak mengalami *overfitting* yang signifikan. Catatan *loss final* pada LSTM adalah 0.0122 (*training*) dan 0.0104 (*validation*), sedangkan CNN-LSTM mencapai 0.0126 dan 0.0106, yang meskipun sedikit lebih tinggi namun masih sangat dekat menunjukkan kedua model memiliki kemampuan pelatihan yang sebanding dalam kondisi dataset yang sama.

### 3.3 Mengevaluasi Model

**Tabel 2.** Perbandingan Evaluasi Model LSTM dan CNN-LSTM

Model	Training Loss	RSME	MAE	MSE	R2	Prediksi
LSTM	0.0122	0.83	0.60	0.68	0.58	5.42
CNN-LSTM	0.0126	0.79	0.63	0.62	0.61	5.66

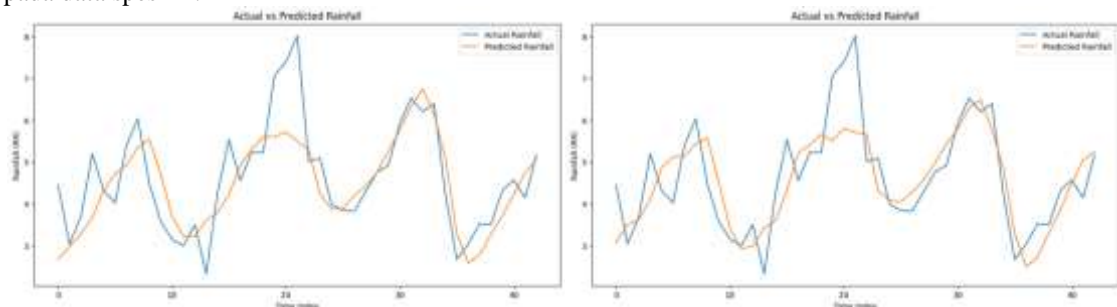
Evaluasi performa pada data uji menggunakan metrik *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *koefisien determinasi* ( $R^2$ ) memberikan gambaran lengkap akurasi dan kemampuan model. Model CNN-LSTM memperoleh RMSE 0.79, MAE 0.63, MSE 0.62, dan  $R^2$  0.61, sementara LSTM memiliki RMSE 0.83, MAE 0.60, MSE 0.68, dan  $R^2$  0.58.

Performa CNN-LSTM yang lebih baik pada RMSE dan MSE sangat penting karena kedua metrik ini menempatkan bobot lebih pada kesalahan besar, yang menjadi krusial dalam prediksi curah hujan yang biasanya memiliki varian ekstrem. Nilai RMSE dan MSE yang lebih rendah menandakan bahwa CNN-LSTM lebih mampu mengurangi kesalahan prediksi yang besar, menghasilkan model yang lebih andal untuk peringatan dini dan mitigasi bencana.

Di sisi lain, MAE LSTM yang sedikit lebih rendah menunjukkan model ini memiliki kesalahan rata-rata yang lebih kecil, yang berarti untuk prediksi harian tanpa fluktuasi ekstrim, LSTM sederhana masih cukup baik dan sedikit lebih presisi. Ini menunjukkan kedua model memiliki kekuatan pada aspek evaluasi berbeda, tetapi secara total performa CNN-LSTM unggul karena menangani variasi temporal dan spasial dengan lebih baik.

Nilai  $R^2$  yang lebih tinggi pada CNN-LSTM juga berarti model ini mampu menangkap proporsi variansi data curah hujan dengan lebih akurat, menandakan bahwa prediksi yang diberikan lebih konsisten dengan pola data asli. Hal ini mengonfirmasi bahwa pemanfaatan lapisan konvolusi dalam CNN-LSTM memberikan keunggulan pada pemahaman pola spasial fitur meteorologi yang kompleks dan bervariasi dalam jangka waktu yang bersangkutan.

CNN-LSTM unggul pada RMSE dan MSE karena kemampuan CNN mengekstrak pola spasial, mengurangi kesalahan besar pada fluktuasi ekstrem curah hujan tropis. MAE LSTM lebih baik untuk kesalahan rata-rata, tapi secara keseluruhan CNN-LSTM lebih andal. Implikasi dari perbedaan RMSE (0.04) berarti prediksi banjir lebih akurat, mengurangi risiko *over/under-estimasi* di lapangan. Dibandingkan studi tropis, hasil ini lebih baik daripada CNN-LSTM untuk typhoon rainfall (RMSE ~1.0), berkat dataset panjang. Keterbatasan: CNN-LSTM memerlukan komputasi lebih tinggi, potensial *overfitting* pada data spesifik.



**Gambar 3.** Perbandingan aktual dan prediksi model LSTM vs CNN-LSTM

Visualisasi hasil prediksi kedua model terhadap data aktual memperlihatkan bahwa CNN-LSTM mampu mengikuti pola puncak dan lembah curah hujan lebih tepat dibanding LSTM yang terkadang melembutkan fluktuasi ekstrim. Hal tersebut mendukung temuan metrik evaluasi yang menyatakan bahwa CNN-LSTM lebih unggul dalam menangani kompleksitas



spasial-temporal. Kurva ini sangat penting untuk analisis aspek praktikal dan penerapan nyata dalam monitoring curah hujan.

Meskipun demikian, perbedaan hasil tidak terlalu besar sehingga model sederhana LSTM tetap memiliki nilai pragmatis terutama bagi kasus dengan keterbatasan sumber daya komputasi. Model CNN-LSTM yang lebih kompleks mengharuskan komputasi lebih tinggi dan mungkin membutuhkan optimasi tambahan agar efisiensi pelatihan meningkat. Penyesuaian hyperparameter dan teknik augmentasi data juga bisa dilakukan untuk mengatasi hal ini di penelitian selanjutnya.

Secara keseluruhan, hasil ini menunjukkan bahwa model *hybrid* CNN-LSTM sangat potensial untuk diaplikasikan pada sistem prediksi curah hujan wilayah dengan karakteristik spasial dan temporal kompleks seperti Kota Tangerang Selatan. Rekomendasi dimasukkan pada sistem peringatan dini bencana dan perencanaan pengelolaan sumber daya air yang memerlukan ketelitian tinggi terutama pada kejadian ekstrim curah hujan. Hasil evaluasi performa kedua model diperkuat oleh temuan dalam penelitian lain yang melaporkan keunggulan CNN-LSTM dalam menangani variasi spasial dan temporal yang kompleks pada data curah hujan wilayah tropis [22][23]. Responsivitas model CNN-LSTM terhadap pola tren historis curah hujan juga sejalan dengan studi yang menunjukkan performa model yang adaptif terhadap perubahan lokal dan musiman [24][25].

Penelitian selanjutnya disarankan untuk menambahkan variabel iklim seperti tekanan udara, intensitas angin, dan faktor eksternal lain yang bisa memperkaya model dan meningkatkan performa prediksi. Selain itu, eksplorasi metode lanjutan seperti attention mechanism, transfer learning, dan metode ensemble juga dapat memperbaiki akurasi dan generalisasi model dalam berbagai kondisi.

#### 4. KESIMPULAN

Penelitian ini membuktikan bahwa model CNN-LSTM menunjukkan kinerja yang lebih unggul dibandingkan model LSTM sederhana dalam memprediksi curah hujan di Kota Tangerang Selatan, dilihat dari metrik evaluasi RMSE, MSE, dan  $R^2$ . Model CNN-LSTM memiliki nilai RMSE dan MSE yang lebih rendah, yang menandakan kesalahan prediksi rata-rata yang lebih kecil dan lebih sedikit kesalahan besar. Selain itu, nilai  $R^2$  yang lebih tinggi pada model CNN-LSTM menunjukkan bahwa model ini mampu menjelaskan variasi data curah hujan dengan lebih baik, sehingga prediksi yang dihasilkan menjadi lebih akurat dan andal secara keseluruhan.

Meskipun model LSTM sederhana tercatat memiliki nilai MAE yang sedikit lebih baik, yang berarti kesalahan absolut rata-rata sedikit lebih kecil, bobot yang lebih besar yang diberikan oleh metrik RMSE pada kesalahan besar membuat CNN-LSTM lebih sesuai untuk aplikasi prediksi curah hujan. Dalam konteks pemodelan hidrometeorologi, RMSE sering dianggap sebagai metrik kunci karena kesalahan besar dalam prediksi memiliki dampak yang signifikan pada pengambilan keputusan, seperti dalam mitigasi bencana banjir dan manajemen sumber daya air. Keunggulan CNN-LSTM pada RMSE, MSE, dan  $R^2$  mengindikasikan bahwa model ini lebih efektif dalam mengidentifikasi pola spasial dan temporal yang kompleks dan mampu menangani fluktuasi ekstrim dalam data curah hujan.

Penelitian ini membuktikan CNN-LSTM unggul atas LSTM untuk prediksi curah hujan di Tangerang Selatan, dengan RMSE lebih rendah dan  $R^2$  lebih tinggi, menekankan kontribusi akademis pada penerapan hybrid untuk dataset tropis jangka panjang. Variabel iklim terbatas, uji coba hanya satu kota, potensial bias normalitas pada uji t.

Rekomendasi dari penelitian ini adalah penggunaan model hybrid CNN-LSTM sebagai pilihan utama dalam prediksi cuaca berbasis data urutan dengan karakteristik spasial dan temporal yang kompleks seperti yang ditemukan di Kota Tangerang Selatan dan juga mengkonfirmasi bahwa model hybrid CNN-LSTM lebih sesuai untuk prediksi curah hujan dengan karakteristik spasial dan temporal kompleks, mirip dengan temuan pada studi terkini di Indonesia dan wilayah tropis. Namun, ada ruang untuk pengembangan lebih lanjut, terutama terkait dengan integrasi variabel iklim tambahan dan optimasi teknik pelatihan untuk meningkatkan kinerja model. Penelitian lanjutan dapat mengeksplorasi teknik seperti attention mechanism, augmentasi data, dan tuning parameter yang lebih mendalam agar presisi prediksi dapat meningkat dan hasilnya semakin aplikatif untuk kebutuhan sistem peringatan dini cuaca dan pengelolaan risiko bencana. Serta dapat pula integrasikan transformer-based models atau variabel tambahan seperti tekanan udara untuk meningkatkan generalisasi.

#### REFERENCES

- [1] D. A. Setiyowati and M. Ariska, "Analisis Pola Curah Hujan Di Pulau Jawa Dengan Menggunakan Empirical Orthogonal Function (EOF)," *J. Pendidik. Fis. dan Sains*, vol. 7, no. 2, pp. 120–128, Sep. 2024, doi: 10.52188/jpfs.v7i2.788.
- [2] L. A. Al-Samraie *et al.*, "Deep Learning Models Based on CNN, RNN, and LSTM for Rainfall Forecasting: Jordan as a Case Study," *Math. Model. Eng. Probl.*, vol. 12, no. 7, pp. 2456–2466, Jul. 2025, doi: 10.18280/mmep.120724.
- [3] A. Azi and K. Kusriani, "Rainfall Prediction in Jayapura City Area Using Long Short-Term Memory," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 7, no. 2, pp. 433–439, Apr. 2025, doi: 10.47709/cnahpc.v7i2.5506.
- [4] A. M. L. Harefa, R. Antoni, A. I. Sitepu, Y. F. Limbong, and M. S. Novelan, "Enhanced Rainfall Forecasting Through Deep Learning Optimization Using Long Short-Term Memory Networks," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 274–284, May 2025, doi: 10.31004/riggs.v4i2.487.





- [5] F. I. Arassah, K. Sadik, B. Sartono, and P. Sofan, "Optimizing Machine Learning for Daily Rainfall Prediction in Bogor: A Statistical Downscaling Approach," *Eduvest - J. Univers. Stud.*, vol. 5, no. 6, pp. 7006–7018, Jun. 2025, doi: 10.59188/eduvest.v5i6.51307.
- [6] Z. N. Ruslana and E. Zuliarso, "Rainfall Forecasting Using SSA-Based Hybrid Models with LSSVR and LSTM for Disaster Mitigation," vol. 6, no. 4, pp. 2079–2106, 2025.
- [7] N. A. Fadhilah *et al.*, "Long Short-Term Memory as a Rainfall Forecasting Model for Bogor City in 2025-2026," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 333–340, Mar. 2025, doi: 10.30871/jaic.v9i2.9068.
- [8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
- [9] S. Samantaray, A. Sahoo, and D. P. Satapathy, "Rainfall Prediction using Hybrid CNN-LSTM approach: A case study in the Boudh district, Odisha, India." Mar. 2025. doi: 10.5194/egusphere-egu25-1542.
- [10] M. F. Putranto and R. Munir, "Deep Learning Approach for Heavy Rainfall Prediction Using Himawari-8 and RDCA Data," *Proc. - 2023 10th Int. Conf. Comput. Control. Informatics its Appl. Explor. Power Data Leveraging Inf. to Drive Digit. Innov. IC3INA 2023*, pp. 424–429, 2023, doi: 10.1109/IC3INA60834.2023.10285744.
- [11] B. K. Cahyono *et al.*, "Leveraging machine learning and open accessed remote sensing data for precise rainfall forecasting," *Commun. Sci. Technol.*, vol. 10, no. 1, pp. 135–147, Jul. 2025, doi: 10.21924/cst.10.1.2025.1638.
- [12] B. Ladjal *et al.*, "Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria," *Sci. Rep.*, vol. 15, no. 1, p. 15404, May 2025, doi: 10.1038/s41598-025-94239-z.
- [13] E. Pitaloka and S. S. Tegar Bagus Agung Hartanto, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *J. Sist. Inf. Bisnis*, vol. 14, no. 1, pp. 62–76, 2024, doi: 10.21456/vol14iss1pp62-76.
- [14] E. Lee *et al.*, "Characterization of naturally occurring radioactive material dynamics in community water systems using groundwater from Ganghwa Island, Republic of Korea," *J. Hydrol.*, vol. 628, p. 130512, Jan. 2024, doi: 10.1016/j.jhydrol.2023.130512.
- [15] Q. Guo, Z. He, and Z. Wang, "Monthly climate prediction using deep convolutional neural network and long short-term memory," *Sci. Rep.*, vol. 14, no. 1, pp. 1–18, 2024, doi: 10.1038/s41598-024-68906-6.
- [16] M. J. Uddin, Y. Li, M. A. Sattar, M. Liu, and N. Yang, "An Improved Cluster-Wise Typhoon Rainfall Forecasting Model Based on Machine Learning and Deep Learning Models Over the Northwestern Pacific Ocean," *J. Geophys. Res. Atmos.*, vol. 127, no. 14, Jul. 2022, doi: 10.1029/2022JD036603.
- [17] M. Waqas, U. W. Humphries, and P. T. Hlaing, "Time series trend analysis and forecasting of climate variability using deep learning in Thailand," *Results Eng.*, vol. 24, p. 102997, Dec. 2024, doi: 10.1016/j.rineng.2024.102997.
- [18] J. Derot, N. Sugiura, S. Kim, and S. Kouketsu, "Improved climate time series forecasts by machine learning and statistical models coupled with signature method: A case study with El Niño," *Ecol. Inform.*, vol. 79, p. 102437, Mar. 2024, doi: 10.1016/j.ecoinf.2023.102437.
- [19] K. A. Koparanov, K. K. Georgiev, and V. A. Shterev, "Lookback Period, Epochs and Hidden States Effect on Time Series Prediction Using a LSTM based Neural Network," in *2020 28th National Conference with International Participation (TELECOM)*, IEEE, Oct. 2020, pp. 61–64. doi: 10.1109/TELECOM50385.2020.9299551.
- [20] Nurfalinda, M. Al Fiani, and M. R. Rathomi, "Prediksi Temperatur Maksimum di Kota Tanjungpinang Menggunakan Model CNN-LSTM," *Komputa J. Ilm. Komput. dan Inform.*, vol. 14, no. 1, pp. 129–138, May 2025, doi: 10.34010/komputa.v14i1.15377.
- [21] R. Yuan, "Rainfall prediction based on CNN-LSTM model under sliding window," *Eur. J. Remote Sens.*, vol. 58, no. 1, Dec. 2025, doi: 10.1080/22797254.2025.2540106.
- [22] A. Saleh, M. L. Tan, Z. M. Yaseen, and F. Zhang, "Integrated machine learning models for enhancing tropical rainfall prediction using NASA POWER meteorological data," *J. Water Clim. Chang.*, vol. 15, no. 12, pp. 6022–6042, Dec. 2024, doi: 10.2166/wcc.2024.719.
- [23] V. Monita, Sevirda Raniprima, and Nanang Cahyadi, "Comparative Analysis of Daily and Weekly Heavy Rain Prediction Using LSTM and Cloud Data," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 4, pp. 833–842, 2024, doi: 10.26555/jiteki.v10i4.30374.
- [24] A. Rasheeda Satheesh, P. Knippertz, and A. H. Fink, "Machine Learning Models for Daily Rainfall Forecasting in Northern Tropical Africa Using Tropical Wave Predictors," *Weather Forecast.*, vol. 40, no. 10, pp. 1895–1916, Oct. 2025, doi: 10.1175/WAF-D-24-0192.1.
- [25] M. Sinambela, S. Humaidi, and M. Situmorang, "Application of Convolutional Neural Networks ( CNN ) for Optimizing Route Changes Based on Dynamic Weather Conditions and Travel Time," pp. 183–197.
- [26] S. Salma and A. Ashwitha, "Hybrid Cnn-Lstm Model: Rainfall Analysis and Prediction for Karnataka Region," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 22, pp. 6715–6716, 2022.

