

Prediksi Solar Irradiance Berbasis Citra Langit Menggunakan Multi-Stage Cross-Modal Attention Transformer

Jeffry Antoni
Prof. Dr.Ir. Rinaldi, M.T

Program Studi Magister Informatika
Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung
Bandung, Indonesia
Email: 23524020@std.stei.itb.ac.id

Abstrak—Prediksi *solar irradiance* jangka pendek berbasis citra langit menghadapi dua tantangan utama: fluktuasi tinggi pada *Global Horizontal Irradiance* (GHI) akibat dinamika awan konvektif pada iklim tropis, serta keterbatasan kapasitas representasi pada mekanisme fusi satu tahap yang hanya mempelajari interaksi lintas-modalitas pada satu titik dalam jaringan. Paper ini mengusulkan *Multi-Stage Cross-Modal Attention Transformer* untuk prediksi GHI pada horizon 30 menit, yang mengintegrasikan modalitas citra langit dan deret waktu secara bertahap melalui mekanisme *multi-stage cross-modal attention*. Model menggunakan *encoder Informer* untuk jalur deret waktu dan *Vision Transformer* (ViT) yang beroperasi pada peta *optical flow Farneback* untuk jalur citra, dengan tambahan fitur suhu udara dan kelembaban relatif. Menggunakan 23.869 citra langit dan data sensor tersinkronisasi dari PLTS PLN Puslitbang Jakarta (Februari 2023–Januari 2024) yang menghasilkan 18.699 sequence, konfigurasi usulan 4-stage mencapai NRMSE 10,27%, RMSE 82,31 W/m², dan R² 0,799, mengungguli baseline Transformer multimodal state-of-the-art (NRMSE 10,56%), baseline LSTM+CNN (NRMSE 10,59%), serta model ablasi modalitas tunggal (*InformerOnly* 10,37%, *ViTOnly* 18,32%). Studi sensitivitas mengungkap pola U-shape antara jumlah *stage fusion* dan NRMSE, dengan 4-stage sebagai konfigurasi optimal, serta menunjukkan bahwa peningkatan resolusi citra input di atas 100×100 piksel tidak meningkatkan akurasi namun secara signifikan meningkatkan waktu inferensi (1,19 ms/sampel pada konfigurasi optimal).

Kata Kunci—*cross-modal attention, multi-stage fusion, optical flow, prediksi solar irradiance, Vision Transformer, Informer, citra langit, GHI*

I. PENDAHULUAN

Pembangkit Listrik Tenaga Surya (PLTS) merupakan salah satu pilar utama transisi energi terbarukan di Indonesia. Kapasitas terpasang PLTS nasional mencapai 909,4 MW pada tahun 2024, dengan target 1.615 MW pada tahun 2025, dan RUPTL PLN 2025–2034 mengalokasikan porsi signifikan sebesar 69.512 MW dari rencana pembangkit dan sistem penyimpanan energi untuk sumber energi terbarukan [8]. Namun, berbeda dengan pembangkit konvensional, daya keluaran PLTS sangat dipengaruhi oleh solar irradiance, yang diukur sebagai *Global Horizontal Irradiance* (GHI, W/m²), yang dapat berubah secara cepat akibat faktor *intermittency*. Awan konvektif yang bergerak cepat, khususnya, menyebabkan fluktuasi dalam skala menit yang sulit diantisipasi pada sistem kelistrikan seperti Jawa-Madura-Bali [1], sementara kondisi atmosfer lokal seperti suhu dan kelembaban turut memodulasi pelemahan radiasi yang masuk.

Dua keterbatasan melatarbelakangi penelitian ini. Pertama, pendekatan Transformer multimodal yang ada untuk prediksi GHI berbasis citra langit umumnya melakukan fusi representasi visual dan temporal hanya pada satu tahap dalam jaringan, sehingga membatasi kedalaman

interaksi lintas-modalitas yang dapat dipelajari. Kedua, banyak pipeline prediksi deret waktu mengabaikan variabel meteorologi eksplisit seperti suhu dan kelembaban, hanya mengandalkan GHI historis dan *clear-sky GHI*.

Berdasarkan bukti dari domain deteksi objek 3D bahwa fusi lintas-modalitas bertahap menghasilkan representasi gabungan yang lebih kaya dibandingkan fusi satu tahap, paper ini mengusulkan *Multi-Stage Cross-Modal Attention Transformer* yang mempelajari interaksi visual-temporal secara bertahap melalui beberapa stage, dengan skip connection yang mengakumulasi informasi dari stage-stage sebelumnya. Jalur deret waktu diperkaya lebih lanjut dengan fitur suhu udara dan kelembaban relatif.

Kontribusi paper ini adalah: (1) mekanisme *multi-stage cross-modal attention* untuk prediksi GHI berbasis citra langit, yang mengembangkan baseline Transformer multimodal satu tahap state-of-the-art; (2) penambahan fitur meteorologi (suhu, kelembaban) pada jalur deret waktu; (3) studi sensitivitas terhadap jumlah *stage fusion* dan resolusi citra input; dan (4) studi ablasi yang mengkuantifikasi kontribusi individual modalitas citra dan deret waktu, divalidasi pada dataset PLTS operasional tropis dari Jakarta, Indonesia.

Secara spesifik, paper ini menjawab rumusan masalah berikut: (RM1) Bagaimana merancang Multi-Stage Cross-Modal Attention Transformer untuk memfusiikan citra langit dan data deret waktu pada prediksi GHI jangka pendek? (RM2) Berapa jumlah stage fusion dan resolusi citra input yang memberikan trade-off terbaik antara akurasi prediksi dan efisiensi inferensi? (RM3) Berapa besar kontribusi individual modalitas citra langit, modalitas deret waktu, dan fitur meteorologi tambahan terhadap akurasi prediksi keseluruhan? Hipotesis yang mendasari adalah bahwa model dengan fusi lintas-modalitas bertahap akan menghasilkan NRMSE lebih rendah dibandingkan model dengan fusi satu tahap, bahwa terdapat jumlah stage optimal di mana penambahan stage berlebihan akan menurunkan performa akibat kompleksitas jaringan yang meningkat, dan bahwa penambahan fitur suhu dan kelembaban akan lebih lanjut meningkatkan akurasi.

Sisa paper ini disusun sebagai berikut. Bagian II meninjau penelitian terkait mengenai prediksi irradiance berbasis citra langit dan fusi lintas-modalitas. Bagian III menjelaskan dataset dan pipeline preprocessing. Bagian IV menyajikan arsitektur usulan. Bagian V merinci rancangan eksperimen, dan Bagian VI melaporkan serta membahas hasil. Bagian VII menyimpulkan paper ini.

II. PENELITIAN TERKAIT

A. Prediksi Irradiance Berbasis Citra Langit

Metode prediksi irradiance berbasis citra langit umumnya menggabungkan image encoder berbasis konvolusi atau Transformer dengan model sekuens untuk irradiance historis. Dibandingkan dengan pendekatan berbasis satelit atau Numerical Weather Prediction (NWP), all-sky imager berbasis darat menawarkan resolusi spasial dan temporal yang lebih tinggi, sesuai untuk horizon nowcasting sangat pendek (beberapa menit hingga sekitar satu jam), di mana awan konvektif yang berkembang cepat mendominasi variabilitas irradiance [13], [15], [9]. Optical flow umum digunakan sebagai tahap pra-pemrosesan untuk mengekstraksi pergerakan awan secara eksplisit, karena awan bersifat difus tanpa fitur sudut atau tepi yang jelas, sehingga estimasi dense optical flow lebih sesuai dibandingkan pendekatan sparse feature-tracking untuk menangkap pergerakan di seluruh area langit yang terlihat [12].

B. Sequence Transformer dan Vision Transformer

Untuk pemodelan temporal, arsitektur Informer memperluas Transformer standar [18] dengan ProbSparse self-attention dan operasi distilling untuk memodelkan sekuens deret waktu panjang secara efisien dengan biaya komputasi lebih rendah [20], dan telah diadaptasi untuk prediksi irradiance mengingat kebutuhan memproses GHI

historis, clear-sky GHI, dan variabel meteorologi tambahan. Pada sisi visual, Vision Transformer (ViT) membagi citra menjadi patch berukuran tetap yang tidak saling tumpang tindih, melakukan embedding linear pada setiap patch, dan menerapkan self-attention pada sekuens token patch yang dihasilkan [4], memberikan mekanisme alami untuk menyelaraskan token visual dengan representasi berbasis sekuens yang dihasilkan encoder deret waktu untuk cross-modal attention selanjutnya.

C. Fusi Lintas-Modalitas

Baseline state-of-the-art yang representatif adalah Transformer multimodal yang memfusiikan encoder deret waktu Informer dan encoder optical-flow berbasis ViT melalui satu blok cross-modality attention [11]. Meskipun efektif, desain satu tahap ini membatasi sejauh mana representasi visual dan temporal dapat saling menyempurnakan secara iteratif. Secara terpisah, pada domain deteksi objek 3D, fusi lintas-modalitas bertahap antara fitur LiDAR dan kamera terbukti menghasilkan representasi gabungan yang lebih kaya dibandingkan fusi pada satu titik [7], yang melatarbelakangi desain fusi bertahap yang diadopsi pada paper ini. Hal ini konsisten dengan kajian yang lebih luas mengenai strategi alignment dan fusi multimodal [10]. Praktik evaluasi pada penelitian prediksi solar irradiance terkini cenderung menggunakan protokol multi-metrik (ME, MAE, RMSE, NRMSE, R^2) dengan NRMSE sebagai kriteria utama yang bersifat scale-independent untuk membandingkan model antar dataset [5], yang diikuti pada paper ini.

Dari segi rentang akurasi yang dilaporkan, penelitian prediksi GHI berbasis citra langit sebelumnya melaporkan nilai RMSE berkisar antara 49 hingga 177 W/m² tergantung kondisi langit (cerah, sebagian berawan, atau mendung) [14], [16], sementara model prediksi berbasis citra langit untuk horizon 2 menit ke depan dilaporkan mencapai RMSE 83,8 W/m² [19]. Studi lain berbasis deep learning dengan ekstraksi fitur multi-skala [3] dan model prediksi jangka pendek berbasis citra langit [2] melengkapi gambaran kinerja pada rentang tersebut. Angka-angka ini memberikan konteks eksternal untuk menginterpretasikan RMSE yang dicapai pada paper ini untuk horizon yang jauh lebih panjang (30 menit).

D. Gap Penelitian

Model prediksi multimodal berbasis citra langit yang ada umumnya (i) membatasi interaksi lintas-modalitas pada satu titik fusi, dan (ii) mengabaikan kovariat meteorologi eksplisit selain irradiance itu sendiri, mengandalkan jalur citra saja untuk menangkap konteks atmosfer secara implisit, sebagaimana juga dicatat pada sistem prediksi operasional yang digunakan PLN [17]. Selain itu, sebagian besar dataset benchmark publik berasal dari iklim lintang menengah atau

sedang, sementara wilayah tropis seperti Indonesia mengalami tutupan awan konvektif yang persisten dan sangat dinamis, berbeda secara struktural dari rezim awan pada benchmark tersebut. Paper ini mengatasi kedua gap tersebut secara bersamaan melalui desain fusi bertahap dan fitur yang diperkaya, divalidasi pada dataset operasional tropis.

III. DATASET DAN PREPROCESSING

A. Sumber Data

Dataset dikumpulkan di PLTS PLN Puslitbang, Jl. Duren Tiga Raya No. 102, Jakarta Selatan (lintang $-6,25^\circ$, bujur $106,85^\circ$), dari Februari 2023 hingga Januari 2024. Dua sumber sensor digunakan: (i) citra langit dari all-sky imager ASI-16 dengan interval 10 menit, total 23.869 citra, dan (ii) Automatic Weather Station (AWS) yang merekam GHI, suhu udara, dan kelembaban relatif pada resolusi 1 menit, total 441.276 baris.

B. Pembersihan, Sinkronisasi, dan Pembentukan Sequence

Data AWS mentah menunjukkan masalah kualitas data (gap, outlier, dan noise sensor) yang ditangani melalui proses pembersihan sebelum pemodelan. Citra langit dan pembacaan sensor disinkronkan berdasarkan timestamp, dan citra yang tidak memiliki pasangan dieliminasi. Clear-sky GHI diturunkan mengikuti model Ineichen–Perez [6] sebagai sinyal referensi yang merepresentasikan irradiance teoretis di bawah langit tanpa awan, memungkinkan clear-sky index digunakan kemudian untuk menstratifikasi evaluasi berdasarkan kondisi langit (cerah, berawan, mendung).

Dense optical flow dihitung menggunakan metode Farneback (cv2.calcOpticalFlowFarneback, dengan $\text{pyr_scale} = 0,5$, $\text{levels} = 3$, $\text{winsize} = 15$, $\text{iterations} = 3$, $\text{poly_n} = 5$, $\text{poly_sigma} = 1,2$) antara frame citra berurutan, menghasilkan peta aliran horizontal dan vertikal (F_x , F_y) yang digabungkan menjadi tensor berukuran $H \times W \times 2$ per pasangan frame. Dari jendela $L_x + 1 = 7$ citra, dihasilkan $L_x = 6$ peta optical flow, sesuai dengan jendela historis 6-langkah (60 menit) yang digunakan pada jalur deret waktu. Horizon prediksi mencakup $L_f = 3$ langkah (+10, +20, +30 menit), dengan paper ini berfokus pada prediksi 30 menit ke depan. Fitur deret waktu (GHI, clear-sky GHI, suhu, kelembaban) dan target prediksi dinormalisasi menggunakan Z-score, dengan scaler yang dilatih hanya pada subset training untuk mencegah kebocoran data.

Setelah pembersihan dan sinkronisasi, dataset menghasilkan 18.699 sequence yang dapat digunakan. Pembagian data berbasis waktu (time-based split) digunakan untuk mencegah temporal leakage, menghasilkan pembagian seperti pada Tabel I. Split ini memusatkan periode data uji ke akhir jendela observasi (Oktober 2023–Januari 2024), yang

bertepatan dengan musim hujan dan dibahas lebih lanjut pada Bagian VI-F.

TABEL I PEMBAGIAN DATASET (BERBASIS WAKTU)

Split	Sequence	Persentase
Train	10.384	55,5%
Validasi	3.902	20,9%
Test	4.413	23,6%
Total	18.699	100%

Seluruh model diimplementasikan dalam Python menggunakan pustaka deep learning standar dan dilatih pada satu workstation berbasis GPU. Pelatihan dan evaluasi untuk setiap skenario menggunakan split data, pipeline preprocessing, dan metrik evaluasi yang identik, hanya berbeda pada faktor arsitektur yang menjadi variabel studi (jumlah stage fusion, resolusi citra, atau ablasi modalitas), untuk memastikan perbandingan yang adil dan dapat direproduksi di seluruh 28 skenario eksperimen.

IV. MODEL USULAN

A. Arsitektur Baseline

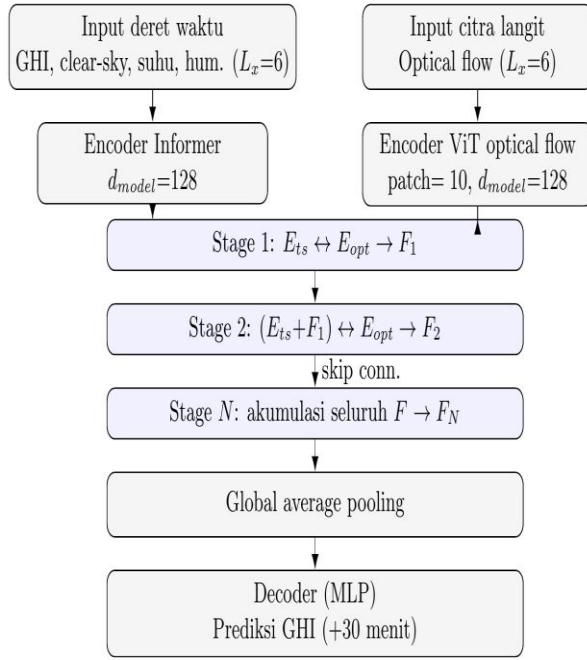
Baseline acuan adalah Transformer multimodal yang terdiri dari: (a) encoder deret waktu Informer yang memproses GHI historis dan clear-sky GHI sepanjang $L_x = 6$ langkah, menghasilkan representasi temporal E_{ts} ; (b) encoder citra ViT yang memproses 6 peta optical flow (diturunkan dari 7 citra langit) menggunakan patch berukuran 10×10 piksel, menghasilkan representasi visual E_{opt} ; dan (c) satu blok cross-modality attention yang memfusi E_{ts} dan E_{opt} sebelum decoding.

B. Multi-Stage Cross-Modal Attention Transformer

Model usulan mempertahankan encoder Informer dan ViT namun menggantikan blok fusi tunggal dengan rangkaian N stage cross-modal attention. Pada setiap stage k , representasi deret waktu diperbarui dengan melakukan attention terhadap representasi visual yang sama, dan hasilnya diteruskan melalui skip connection yang mengakumulasi kontribusi dari seluruh stage sebelumnya:

$$F_k = \text{CrossAttn}(E_{ts} + \sum_{j < k} F_j, E_{opt}), \quad k = 1, \dots, N \quad (1)$$

dengan F_k menyatakan representasi hasil fusi pada stage k . Output stage terakhir F_N , yang mengagregasi informasi terakumulasi dari seluruh stage, dilewatkan melalui global average pooling dan didekode oleh multilayer perceptron (MLP) untuk menghasilkan prediksi GHI pada horizon 30 menit. Gbr. 1 mengilustrasikan alur data secara keseluruhan, dari dua jalur input melalui stage fusion bertumpuk hingga decoder.



Gbr. 1. Alur data Multi-Stage Cross-Modal Attention Transformer yang diusulkan.

Dibandingkan dengan baseline, model usulan juga menambahkan suhu udara dan kelembaban relatif pada jalur deret waktu, memperluas jumlah fitur dari 2 menjadi 4 variabel. Tabel II merangkum hyperparameter model; jumlah stage $N \in \{2,3,4,5,6,7\}$ diperlakukan sebagai variabel sensitivitas utama.

TABEL II HYPERPARAMETER MODEL

Parameter	Nilai
d_model	128
Jumlah head attention	4
Panjang input Lx	6 langkah (60 menit)
Panjang prediksi Lf	3 langkah (+10,+20,+30 menit)
Ukuran citra input	100 × 100 px
Ukuran patch ViT	10 × 10 px (N=100 patch)
Fitur deret waktu	4 (GHI, clear-sky GHI, suhu, hum.)
Stage fusion	2, 3, 4, 5, 6, 7
Fungsi loss	Mean Squared Error (MSE)
Optimizer	Adam ($\beta_1=0.9$; $\beta_2=0.999$)
Learning rate	4×10^{-4}
Batch size	4
Gradient clipping	1,0
Maks. epoch	10
Dropout	0,1

V. RANCANGAN EKSPERIMEN

Seluruh model dievaluasi pada data uji yang identik dan dipisah secara kronologis. Lima metrik digunakan secara bersamaan: Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized RMSE (NRMSE, RMSE dinormalisasi terhadap rentang GHI pada data uji), dan koefisien determinasi (R^2), dengan NRMSE sebagai kriteria utama pemilihan model karena sifatnya yang scale-independent:

$$NRMSE = RMSE / (y_{max} - y_{min}) \times 100\% \quad (2)$$

Peningkatan relatif terhadap baseline dihitung sebagai $(NRMSE_{baseline} - NRMSE_{usulan}) / NRMSE_{baseline} \times 100\%$. Apabila dua model memiliki NRMSE yang berselisih kurang dari 0,1 poin persentase, RMSE digunakan sebagai kriteria pemutus dalam memilih model terbaik.

Tabel III merangkum lingkungan perangkat keras dan perangkat lunak yang digunakan untuk seluruh eksperimen, dijaga identik antar skenario untuk memastikan keterbandingan pengukuran waktu inferensi yang dilaporkan.

TABEL III LINGKUNGAN IMPLEMENTASI

Komponen	Spesifikasi
Bahasa pemrograman	Python
Framework deep learning	PyTorch
Pengolahan citra	OpenCV, PIL (interpolasi bilinear)
Komputasi	Workstation dengan 1 GPU
Protokol evaluasi	Split data uji tetap, metrik identik

Empat model pembanding diimplementasikan: (1) baseline Transformer multimodal satu tahap, dalam dua varian menggunakan 2 fitur (GHI, clear-sky GHI) dan 4 fitur (menambahkan suhu dan kelembaban) untuk mengisolasi kontribusi fitur meteorologi dari kontribusi fusi bertahap; (2) baseline LSTM+CNN menggunakan encoder citra ResNet-18 dan LSTM 2 layer (64 unit), digabungkan melalui konkatenasi; (3) InformerOnly, model ablasi modalitas yang hanya menggunakan jalur deret waktu; dan (4) ViTOnly, yang hanya menggunakan jalur citra. Secara keseluruhan, 28 skenario eksperimen dilakukan, mencakup perbandingan baseline, sensitivitas jumlah stage (2–7 stage), sensitivitas resolusi input (100×100, 160×160, 224×224 px), serta analisis sensitivitas tambahan terhadap strategi pembagian data dan resolusi temporal optical flow.

VI. HASIL DAN PEMBAHASAN

A. Perbandingan Keseluruhan

Tabel IV merangkum performa seluruh model pada data uji. Konfigurasi usulan 4-stage mencapai NRMSE terendah (10,27%) dan RMSE terendah (82,31 W/m²) serta R^2 tertinggi (0,799) di antara seluruh model yang dievaluasi, setara dengan peningkatan relatif NRMSE sebesar 2,75% dibandingkan baseline satu tahap (10,56%). Model ini juga mengungguli baseline LSTM+CNN (10,59%) dan ablasi InformerOnly (10,37%), sementara ablasi ViTOnly menghasilkan performa jauh lebih rendah (18,32%), mengonfirmasi bahwa citra langit saja tidak cukup tanpa konteks temporal.

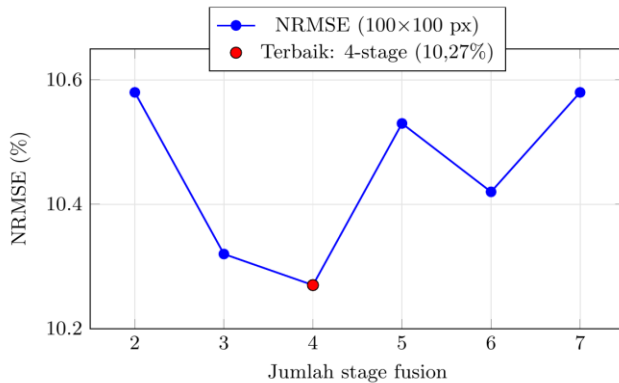
TABEL IV HASIL PADA DATA UJI UNTUK SELURUH MODEL

Model	RMSE (W/m ²)	NRMSE (%)	R^2
Baseline Transformer (2 fitur)	84,66	10,56	0,787
Baseline Transformer (4 fitur)	83,16	10,37	0,790

Model	RMSE (W/m ²)	NRMSE (%)	R ²
LSTM + CNN	84,86	10,59	0,770
InformerOnly (ablasi)	83,09	10,37	0,795
ViTOnly (ablasi)	146,86	18,32	0,360
Usulan, 2-stage	84,77	10,58	0,787
Usulan, 3-stage	82,71	10,32	0,797
Usulan, 4-stage	82,31	10,27	0,799
Usulan, 5-stage	84,43	10,53	0,789
Usulan, 6-stage	83,54	10,42	0,793
Usulan, 7-stage	84,78	10,58	0,787

B. Sensitivitas terhadap Jumlah Stage Fusion

Gbr. 2 memplot NRMSE terhadap jumlah stage fusion pada resolusi 100×100 px (dari Tabel IV), menunjukkan pola U-shape yang jelas: performa membaik dari 2 ke 4 stage dan memburuk setelah 4 stage. Hal ini mengindikasikan bahwa fusi bertahap dalam jumlah moderat memperkaya representasi gabungan, konsisten dengan temuan sebelumnya pada fusi lintas-modalitas bertahap untuk deteksi objek 3D, namun stage yang berlebihan meningkatkan kompleksitas jaringan dan kesulitan optimisasi tanpa peningkatan akurasi yang sepadan, menghasilkan hasil yang menurun (diminishing returns) dan akhirnya negatif. Pola U-shape kualitatif yang sama, namun dengan jumlah stage optimal yang berbeda (2 untuk 160×160 px, 3 untuk 224×224 px), teramati pada dua resolusi lain yang diuji, menunjukkan bahwa kedalaman fusi optimal tidak tetap namun berinteraksi dengan jumlah informasi visual yang tersedia per stage.



Gbr. 2. NRMSE terhadap jumlah stage fusion pada resolusi 100×100 px, menunjukkan pola U-shape dengan titik optimal pada 4 stage.

C. Kontribusi Modalitas

Ablasi InformerOnly (NRMSE 10,37%) menghasilkan performa yang mendekati model usulan terbaik (10,27%), mengindikasikan bahwa modalitas deret waktu dominan pada dataset ini, sementara ablasi ViTOnly (18,32%) mengonfirmasi bahwa informasi visual saja tidak dapat menggantikan konteks temporal. Selisih antara InformerOnly dan model 4-stage lengkap (0,10 poin persentase NRMSE) merepresentasikan kontribusi tambahan dari integrasi informasi citra langit melalui fusi bertahap. Secara terpisah, perluasan fitur deret waktu dari 2 menjadi 4 variabel

(menambahkan suhu dan kelembaban) menurunkan NRMSE sebesar 0,19 poin persentase, mengindikasikan kontribusi positif yang lebih kecil namun konsisten dari fitur meteorologi eksplisit.

D. Pengaruh Resolusi Citra Input

Tabel V merangkum NRMSE terbaik dan waktu inferensi terkait yang dicapai pada setiap resolusi di seluruh konfigurasi stage. Peningkatan resolusi citra input dari 100×100 ke 160×160 dan 224×224 piksel tidak meningkatkan NRMSE pada konfigurasi stage manapun; NRMSE terbaik pada 160×160 (10,31%, 2-stage) dan pada 224×224 (10,43%, 3-stage) keduanya masih berada di atas hasil terbaik pada 100×100 (10,27%, 4-stage). Namun, waktu inferensi meningkat tajam seiring resolusi, dari 1,07–1,41 ms/sampel pada 100×100 menjadi 2,75–4,35 ms/sampel pada 160×160 dan 4,18–9,60 ms/sampel pada 224×224. Hal ini menunjukkan bahwa, untuk dataset dan horizon prediksi ini, konteks visual kasar yang ditangkap pada 100×100 px sudah cukup untuk merepresentasikan pola pergerakan awan yang relevan bagi prediksi GHI, sehingga resolusi lebih tinggi menjadi biaya komputasi yang tidak diperlukan.

TABEL V HASIL TERBAIK PER RESOLUSI CITRA (DI SELURUH JUMLAH STAGE)

Resolusi	NRMSE Terbaik (%)	Stage Terbaik	Rentang Inf. (ms)
100×100	10,27	4	1,07–1,41
160×160	10,31	2	2,75–4,35
224×224	10,43	3	4,18–9,60

E. Performa Antar Kondisi Langit

Data uji distratifikasi berdasarkan clear-sky index (CSI), yaitu rasio GHI aktual terhadap clear-sky GHI, menjadi kondisi Cerah (CSI ≥ 0,45), Berawan (0,10 ≤ CSI < 0,45), dan Mendung (CSI < 0,10), mengikuti ambang batas yang disesuaikan untuk iklim tropis. Sebagaimana ditunjukkan pada Tabel VI, data uji didominasi oleh kondisi Berawan (72,5%), konsisten dengan tutupan awan yang umumnya tinggi pada iklim tropis Jakarta.

TABEL VI PERFORMA MODEL 4-STAGE PER KONDISI LANGIT

Kondisi	Sampel (%)	ME (W/m ²)	RMSE (W/m ²)	NRMSE (%)
Cerah	10,08	+107,17	129,96	16,21
Berawan	72,54	−18,47	89,06	11,11
Mendung	17,38	−25,96	71,80	8,96

Kondisi Mendung menghasilkan RMSE terendah, karena GHI aktual rendah dan relatif stabil, sehingga model lebih mudah menghasilkan prediksi yang akurat. Kondisi Berawan menunjukkan variabilitas error tertinggi, karena rezim transisi ini berkaitan dengan ramp events awan yang tidak terprediksi dan secara inheren paling sulit diprediksi. Kondisi Cerah menunjukkan RMSE tertinggi disertai ME positif

besar (bias under-predict), yang kemungkinan mencerminkan representasi sampel cerah yang rendah pada data latih (26,4%) dibandingkan proporsinya pada data uji, diperparah oleh glare matahari pada citra kamera langit yang menurunkan kualitas optical flow pada kondisi langit cerah.

F. Efisiensi Inferensi

Waktu inferensi untuk model usulan berkisar dari 1,07 ms/sampel (2-stage) hingga 1,41 ms/sampel (7-stage), dengan konfigurasi 4-stage terbaik membutuhkan 1,19 ms/sampel. Angka-angka ini konsisten dengan kebutuhan sistem prediksi operasional real-time berskala menit untuk integrasi jaringan listrik.

G. Implikasi Operasional dan Keterbatasan

Pembagian data latih/uji berbasis waktu menghasilkan pergeseran distribusi musiman, dengan data latih didominasi musim kemarau dan data uji didominasi musim hujan; pergeseran ini diidentifikasi sebagai faktor yang berkontribusi terhadap residual error pada kondisi cuaca tertentu. Eksperimen tambahan menggunakan split interleaved (non-kronologis) menunjukkan penurunan performa keseluruhan dibandingkan split kronologis, mengindikasikan bahwa data uji kronologis — yang terkonsentrasi pada Oktober 2023–Januari 2024 — relatif lebih mudah karena kondisi langit musim hujan yang lebih homogen, sementara split interleaved lebih baik merepresentasikan variasi cuaca sepanjang tahun, termasuk masa transisi musim yang lebih sulit diprediksi. Karena performa pada split interleaved justru menurun alih-alih membaik, pola ini tidak konsisten dengan temporal leakage yang meningkatkan hasil pada split kronologis, dan sebaliknya mengindikasikan bahwa representativitas musiman data uji, bukan kebocoran informasi, merupakan faktor dominan yang menjelaskan perbedaan antara kedua split tersebut.

Selain itu, eksperimen tambahan membandingkan dua cara penurunan optical flow pada periode pelatihan yang sama (Juni–Juli 2023): akumulasi flow dari interval citra 1 menit versus penghitungan flow langsung antar interval 10 menit. Sebagaimana ditunjukkan pada Tabel VII, optical flow interval 10 menit secara konsisten mengungguli varian akumulasi 1 menit, mengindikasikan bahwa granularitas temporal yang lebih halus pada estimasi pergerakan tidak diterjemahkan menjadi peningkatan akurasi prediksi GHI jangka pendek untuk horizon prediksi ini, dan bahwa interval standar imager 10 menit sudah memadai.

TABEL VII PERBANDINGAN RESOLUSI TEMPORAL OPTICAL FLOW (MODEL 4-STAGE)

Skenario	RMSE (W/m ²)	NRMSE (%)	R ²
Optical flow akumulasi 1 menit	96,97	13,24	0,73

Skenario	RMSE (W/m ²)	NRMSE (%)	R ²
Optical flow langsung 10 menit	90,58	12,38	0,76

VII. KESIMPULAN

Paper ini menyajikan Multi-Stage Cross-Modal Attention Transformer untuk prediksi solar irradiance berbasis citra langit pada horizon 30 menit ke depan. Konfigurasi terbaik (4 stage fusion, 4 fitur deret waktu, resolusi input 100×100 px) mencapai NRMSE 10,27%, RMSE 82,31 W/m², dan R² 0,799 pada dataset PLTS operasional tropis dari Jakarta, Indonesia, mengungguli baseline Transformer multimodal satu tahap state-of-the-art sebesar 2,75% dalam NRMSE relatif, serta baseline LSTM+CNN. Studi sensitivitas mengungkap pola U-shape antara jumlah stage fusion dan akurasi, dengan jumlah stage berlebihan menurunkan performa, dan menunjukkan bahwa resolusi citra input yang lebih tinggi meningkatkan biaya inferensi tanpa meningkatkan akurasi. Studi ablasi mengonfirmasi bahwa jalur deret waktu merupakan kontributor dominan terhadap akurasi prediksi, dengan fusi citra langit dan fitur meteorologi memberikan tambahan yang lebih kecil namun konsisten. Waktu inferensi 1,07–1,41 ms per sampel mendukung kelayakan penggunaan operasional secara real-time.

Penelitian selanjutnya mencakup perluasan dataset ke periode yang lebih panjang dan lokasi yang lebih beragam (misalnya lokasi pesisir, pegunungan, atau luar Jawa) untuk meningkatkan generalisasi model pada berbagai zona iklim tropis Indonesia; eksplorasi mekanisme fusi adaptif seperti gated fusion atau adaptive stage weighting; perluasan model untuk horizon prediksi yang lebih panjang atau multi-horizon (60 menit, 2 jam); serta evaluasi performa pada kondisi cuaca ekstrem seperti hujan lebat, kabut tebal, atau ramp events yang intens, yang merupakan skenario kritis secara operasional bagi stabilitas jaringan listrik.

REFERENSI

- [1] AESI, “Sistem Kelistrikan Jamali: Karakteristik dan Tantangan Operasional,” Asosiasi Energi Surya Indonesia, 2021.
- [2] S. Barhmi, A. Idrissi, H. El Ferouali, dan A. Ghennioui, “Short-term solar irradiance forecasting based on sky images,” *Energy Conversion and Management*, vol. 301, p. 117966, 2024.
- [3] X. Chen, Y. Du, E. Lim, H. Wen, dan L. Jiang, “Sky image-based solar irradiance prediction using deep learning with multi-scale feature extraction,” *Renewable Energy*, vol. 215, p. 118982, 2023.
- [4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2021.
- [5] V. Gayathry, D. Kaliyaperumal, dan S. R. Salkuti, “Seasonal solar irradiance forecasting using artificial intelligence

techniques with uncertainty analysis,” *Scientific Reports*, vol. 14, p. 17945, 2024.

- [6] P. Ineichen dan R. Perez, “A new airmass independent formulation for the Linke turbidity coefficient,” *Solar Energy*, vol. 73, no. 3, pp. 151–157, 2002.
- [7] X. Jiang, S. Zhao, R. Shen, W. Li, R. Yang, Z. Chen, dan W. Wang, “MSHP3D: Multi-stage cross-modal fusion based on hybrid perception for indoor 3D object detection,” *Information Fusion*, vol. 112, p. 102591, 2024.
- [8] Kementerian ESDM, “Capaian Kinerja Subsektor EBTKE Tahun 2024,” Direktorat Jenderal Energi Baru, Terbarukan, dan Konservasi Energi, 2024.
- [9] J. Lee dan S. Park, “Nowcasting of solar irradiance components using all-sky imager and meteorological observations,” *Solar Energy*, 2025.
- [10] Y. Li dan H. Tang, “Multimodal alignment and fusion: A survey,” *arXiv preprint arXiv:2411.17040*, 2024.
- [11] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei, dan G. Sun, “A transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting,” *Applied Energy*, vol. 342, p. 121160, 2023.
- [12] A. Mammoli, G. Terren-Serrano, A. Menicucci, dan T. P. Caudell, “Comparative study of algorithms for cloud motion estimation using sky-imaging data,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5137–5142, 2017.
- [13] Y. Nie, A. S. Zamzam, dan A. Dalton, “Open-source ground-based sky image datasets for very short-term solar forecasting: A comprehensive survey,” *arXiv preprint arXiv:2211.14709*, 2022.
- [14] Q. Paletta, G. Arbod, dan J. Lasenby, “Benchmarking of deep learning irradiance forecasting models from sky images - An in-depth analysis,” *Solar Energy*, vol. 224, pp. 504–519, 2021.
- [15] T. Schmidt, J. Kalisch, E. Lorenz, dan D. Heinemann, “Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts,” *Solar Energy*, vol. 156, pp. 1–17, 2017.
- [16] A. Z. Shirazi, M. Bhatt, dan J. Bhatt, “Sky images for short-term solar irradiance forecast: A comparative study of linear machine learning models,” *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3261–3272, 2021.
- [17] SIUJPTL, “Solar Forecasting System,” 2024. [Online]. Tersedia: <https://siujptl.co.id/pembangkit/kamus/solar-forecasting-system>
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [19] J. Zhang, R. Verschae, S. Nobuhara, dan J. F. Lalonde, “Deep photovoltaic nowcasting,” *Solar Energy*, vol. 176, pp. 267–276, 2022.
- [20] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, dan W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” *Proc. AAAI Conf. on Artificial Intelligence*, vol. 35, no. 12, pp. 11106–11115, 2021.