One-Hour-Ahead Mean Radiant Temperature Forecasting in Jabodetabek Using CNN-LSTM and Temporal Convolutional Networks

Novana Sari*[1], Tukiyat[2], Yan Mitha Djaksana[3]
[1,3]Master of Informatics Engineering Program, Universitas Pamulang, Kota Tangerang Selatan, Banten
[2]Research Center for Limnology and Water Resources, National Research and Innovation Agency (BRIN), Indonesia.

Article Info

Article history:

Received Jul 08, 2026
Revised Jul 10, 2026
Accepted Nov 1, 2026

Corresponding Author:

Novana Sari
Master of Informatics Engineering Program,
Universitas Pamulang, Kota Tangerang Selatan, Banten
Email: novana.sari9@gmail.com

ABSTRACT

Mean Radiant Temperature (MRT) represents the combined shortwave and longwave radiant load experienced by a human body, but continuous observations are rarely available across large metropolitan areas. This study developed and compared a Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model and a Temporal Convolutional Network (TCN) for one-hour-ahead MRT forecasting in Jabodetabek. The dataset comprised 210,528 hourly records from 2023–2024 at 12 ERA5 grid points. Five radiation variables and two near-surface thermal variables were used as predictors, while ERA5-HEAT MRT was the target. Each sample contained a 24-hour historical window. Three chronological train-validation-test splits were evaluated: 80:10:10, 70:10:20, and 70:15:15. Both architectures achieved R^2 values above 0.93 in all scenarios. Under the 80:10:10 split, TCN produced the lowest RMSE value of 2.8104 °C, the highest R^2 value of 0.9518, the lowest validation loss value of 5.9592, and a residual bias of −0.4863 °C. CNN-LSTM achieved the lowest MAE value of 1.8874 °C and MAPE value of 5.76% and was more stable when the training proportion decreased. Overall, TCN 80:10:10 was selected as the best configuration, although field validation is required before operational deployment.

Keywords: mean radiant temperature; CNN-LSTM; TCN; ERA5; urban thermal forecasting.

1. INTRODUCTION

Rapid urban development changes surface materials, building density, vegetation cover, and the partitioning of surface energy. These changes contribute to urban heat-island conditions and increase the thermal load experienced by urban residents. In a metropolitan region such as Jabodetabek, the interaction between dense built-up areas, coastal influence, moisture, and strong tropical radiation produces substantial temporal and spatial variability in outdoor heat exposure (Oke et al., 2017; Ren et al., 2023). Reliable short-term thermal information is therefore important for heat-risk awareness, outdoor activity planning, and climate-responsive urban management.

Air temperature alone does not fully describe outdoor thermal exposure. Mean Radiant Temperature (MRT) is the equivalent uniform temperature of a radiative environment that would produce the same net radiant heat exchange as the actual surroundings. It integrates direct and diffuse solar radiation, reflected shortwave radiation, atmospheric longwave radiation, and thermal emission from surrounding surfaces (Matzarakis et al., 2010; Thorsson et al., 2007). Because radiant exchange can change rapidly with solar position, cloudiness, shading, and surface properties, MRT often exhibits a stronger diurnal amplitude than near-surface air temperature.

Direct MRT observation remains difficult at metropolitan scale. Globe thermometers or multi-directional radiometers require careful calibration, appropriate wind correction, and dense deployment to represent heterogeneous urban settings. Field experiments have shown that radiant heat may vary considerably over short distances and short time intervals (Aviv et al., 2021). Physically based tools such as SOLWEIG can reproduce three-dimensional radiation fluxes and shading, but they require detailed urban geometry, vegetation, and surface information and become data-intensive when applied over broad domains (Lindberg et al., 2008).

Atmospheric reanalysis offers a complementary source of temporally continuous and spatially consistent information. ERA5 provides hourly global atmospheric and land-surface variables, including surface solar and thermal radiation and near-surface temperature and humidity (Hersbach et al., 2020). ERA5-HEAT derives human thermal-comfort variables, including MRT, from ERA5 radiation and atmospheric conditions (Di Napoli et al., 2021). Reanalysis-based thermal products have also supported global heat-risk indicators, demonstrating the value of gridded radiation fields when direct observations are unavailable (Brimicombe et al., 2023).

Forecasting MRT from multivariate radiation and thermal histories is a nonlinear sequence-learning problem. Deep neural networks can learn hierarchical representations from large datasets and model interactions that are difficult to express through fixed empirical equations (LeCun et al., 2015). A CNN-LSTM architecture first applies one-dimensional convolution to identify local temporal-feature patterns and subsequently uses Long Short-Term Memory units to retain relevant sequential information. Hybrid convolutional-recurrent networks have produced competitive results in multivariate forecasting because they combine local feature extraction with temporal memory (Zhou et al., 2022).

Temporal Convolutional Networks (TCNs) provide a different mechanism. They use causal convolutions, increasing dilation factors, and residual connections to obtain a long receptive field without recurrent processing. Generic sequence-model comparisons and forecasting applications have shown that TCNs can match or exceed recurrent networks while maintaining stable gradient flow (Bai et al., 2018; Lara-Benítez et al., 2020). Weather-forecasting research has further demonstrated that TCNs can effectively represent nonlinear temporal relationships among multiple meteorological variables (Hewage et al., 2020).

Deep learning has been used to accelerate spatial MRT modelling; for example, a convolutional encoder-decoder reproduced long-term SOLWEIG outputs at pedestrian scale with relatively low error (Briegel et al., 2023). Nevertheless, the literature remains limited regarding direct comparison of CNN-LSTM and TCN for one-hour-ahead MRT prediction in a tropical metropolitan region using radiation and thermal variables from ERA5. Most MRT studies emphasize measurement, physical simulation, spatial mapping, or thermal-index construction rather than controlled temporal forecasting across multiple chronological splits.

This study addresses that gap by comparing CNN-LSTM and TCN using the same 24-hour input history, seven ERA5 predictors, identical training controls, and three chronological data-partition scenarios. The contribution is threefold: first, it establishes a reproducible sequence-processing pipeline that prevents leakage between training, validation, and testing periods; second, it evaluates the sensitivity of both architectures to changes in data allocation; and third, it extends aggregate metrics with training-loss, residual-bias, and grid-level error analyses. The research questions concern whether both models can predict MRT one hour ahead, how split composition affects their generalization, and which configuration provides the most balanced statistical performance.

2. RESEARCH METHOD

2.1. DATA AND STUDY AREA

The research used a quantitative comparative-experiment design. Hourly ERA5 and ERA5-HEAT data covered 1 January 2023 to 31 December 2024 over 6.00°–6.50° S and 106.50°–107.25° E. With a 0.25° × 0.25° grid, the domain comprised three latitude positions and four longitude positions, producing 12 grid points. Each grid contained 17,544 hourly observations, resulting in 210,528 initial records. Figure 1 summarizes the analytical workflow, while Figure 2 shows the spatial sampling arrangement.

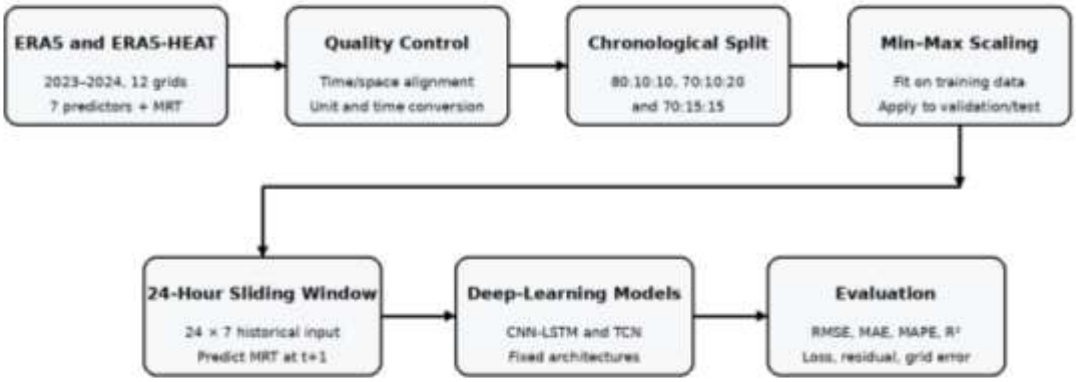


Figure 1. Flowchart Research

Five accumulated ERA5 radiation variables were used: surface net solar radiation (ssr), surface net thermal radiation (str), surface solar radiation downwards (ssrd), surface thermal radiation downwards (strd), and total sky direct solar radiation at the surface (fdir). Two near-surface variables, 2 m air temperature (t2m) and 2 m dew-point temperature (d2m), provided additional thermal and moisture information. MRT from ERA5-HEAT served as the prediction target. The target should be interpreted as a reanalysis-derived gridded estimate rather than direct pedestrian-level observation.

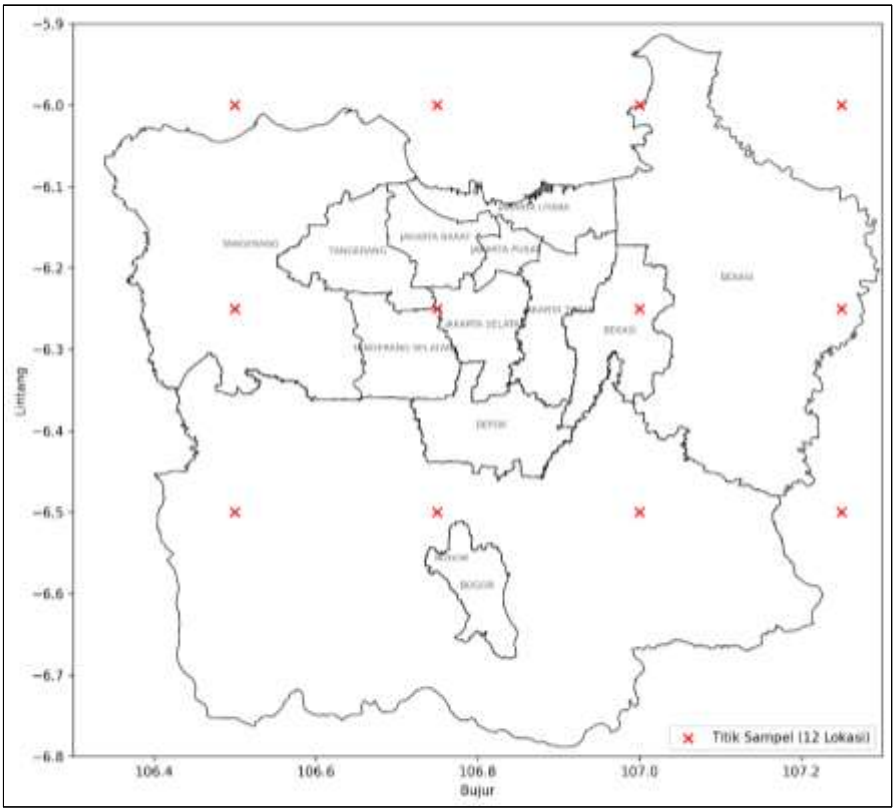


Figure 2. Spatial distribution of the 12 grid points in the Jabodetabek study domain

2.2. DATA PREPARATION AND SEQUNTIAL SAMPLING

Table 1. Sequential samples generated after the 24-hour sliding window

Split	Training samples	Validation samples	Testing samples
80:10:10	168,132	20,760	20,772
70:10:20	147,072	20,772	41,820
70:15:15	147,072	31,296	31,296

2.3. CNN-LSTM AND TCN ARCHITECTURE

The CNN-LSTM received an input tensor of 24 time steps and seven features. A Conv1D layer with 64 filters, kernel size 3, valid padding, and ReLU activation extracted local patterns. Its output was passed to an LSTM layer containing 50 units. LSTM gates regulate memory retention and forgetting, allowing the network to preserve useful temporal information while limiting vanishing-gradient effects (Hochreiter & Schmidhuber, 1997). A linear one-neuron dense layer generated the MRT forecast. The architecture contained 24,459 trainable parameters.

The TCN consisted of four residual blocks with dilation factors 1, 2, 4, and 8. Each block contained two causal Conv1D layers with 64 filters and kernel size 3, followed by batch normalization, ReLU activation, and 0.2 dropout. A 1×1 convolution aligned the residual shortcut in the first block. After the last block, a flatten layer and a linear dense neuron produced the forecast. The calculated receptive field was 61 steps, exceeding the 24-hour input and therefore covering the complete historical window. The model had 91,969 parameters, including 90,945 trainable parameters.

Both models used the Adam optimizer, a learning rate of 0.001, Mean Squared Error as the loss function, a batch size of 32, and a maximum of 100 epochs. Early stopping monitored validation loss with patience of ten epochs, restored the best weights, and worked together with model checkpointing. The random seed was fixed at 42, and shuffle was disabled. Architecture settings were unchanged across split scenarios so that performance differences mainly reflected data allocation rather than model redesign.

Table 2. Sequential samples generated after the 24-hour sliding window

Component	CNN-LSTM	TCN
Input	24 time steps × 7 features	24 time steps × 7 features
Temporal extractor	Conv1D: 64 filters, kernel 3	Four causal residual blocks
Sequence mechanism	LSTM: 50 units	Dilations 1, 2, 4, and 8
Regularization	Early stopping	Batch normalization, dropout 0.2, early stopping
Output	One linear neuron	Flatten + one linear neuron
Parameters	24,459 trainable	91,969 total; 90,945 trainable
Training controls	Adam, learning rate 0.001, MSE, batch 32, maximum 100 epochs	Same

2.4. EVALUATION MODEL

Test-set performance was measured using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). RMSE emphasizes relatively large errors, MAE represents the average absolute deviation, MAPE expresses relative error, and R^2 measures the fraction of target variation reproduced by the model. Because individual metrics emphasize different aspects of performance, they were interpreted jointly rather than used in isolation (Chicco et al., 2021; Hyndman & Koehler, 2006).

Training and validation loss were inspected to determine convergence and overfitting. Residuals were defined as actual MRT minus predicted MRT; therefore, a negative mean residual indicated average overprediction. Scatterplots, temporal trajectories, and hourly climatological profiles were examined to determine whether the models reproduced the daily cycle and extreme changes. Finally, RMSE and residual bias were calculated separately for each grid to assess whether aggregate performance concealed spatially systematic error.

3. RESULT AND ANALYSIS

3.1. DATA CHARACTERISTICS

The quality-controlled dataset contained all 210,528 expected observations. MRT ranged from 9.68 °C to 62.91 °C, with a mean of 33.09 °C and a standard deviation of 13.42 °C. The shortwave predictors showed strongly skewed distributions because nighttime values approached zero while clear daytime conditions produced values near or above 1,000 W m^{-2}. In comparison, t2m and d2m changed more gradually, with means of 27.41 °C and 23.40 °C, respectively.

The diurnal signal was substantially stronger than the month-to-month mean variation. MRT increased rapidly after sunrise, reached an average maximum of 51.10 °C around 11:00 local time, and declined through the afternoon. The lowest hourly average, 20.56 °C, occurred around 05:00. Monthly mean MRT was highest in October (34.80 °C) and lowest in February (31.67 °C). Figure 3 presents the complete month–hour structure as a heatmap.

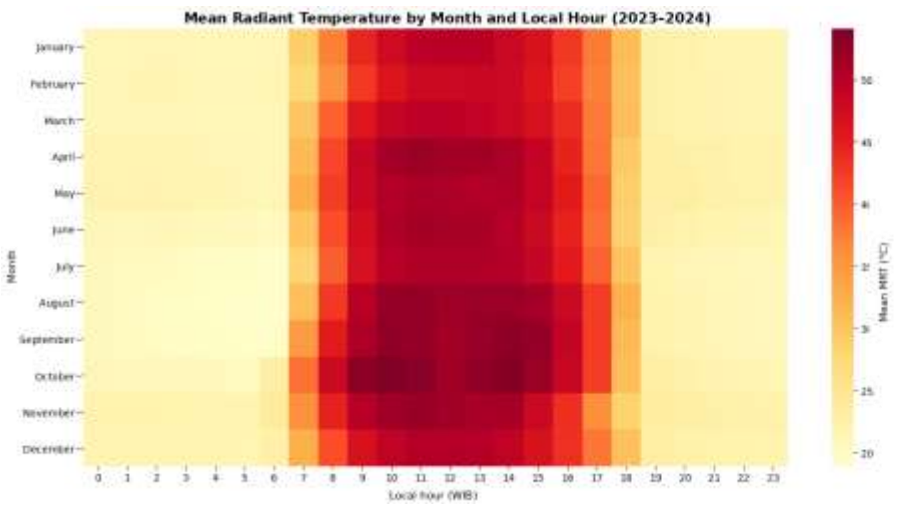


Figure 3. Month-hour heatmap of average MRT during 2023-2024

Correlation analysis showed that downward and net shortwave radiation were the strongest linear predictors. The Pearson coefficients between MRT and ssrd, ssr, and fdir were 0.926, 0.923, and 0.867, respectively. Air temperature also showed a strong positive relationship (r = 0.781). Surface net thermal radiation was negatively correlated with MRT (r = −0.701), whereas strd had a weaker positive association and d2m had only a weak direct linear relationship. The complete predictor set was retained because deep networks can learn nonlinear interactions and lagged information not represented by bivariate correlation.

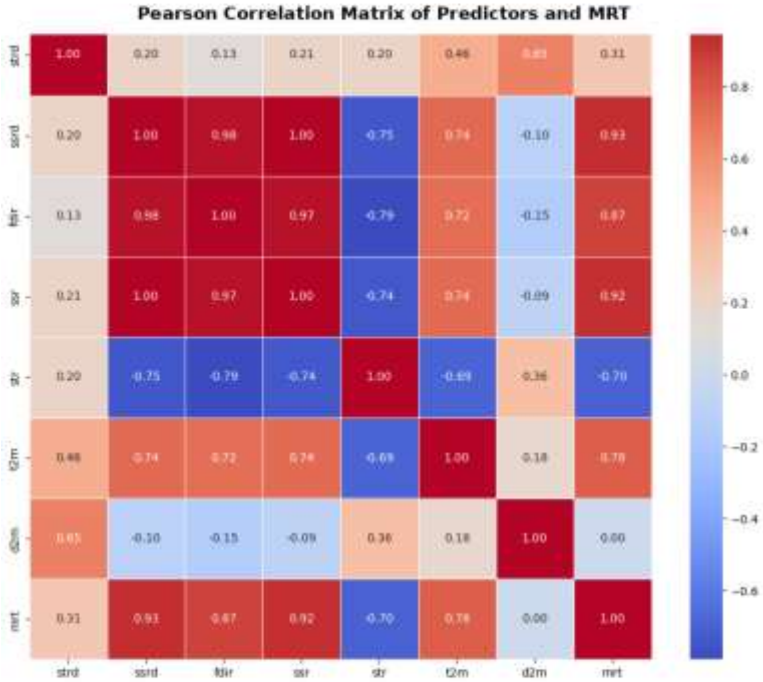
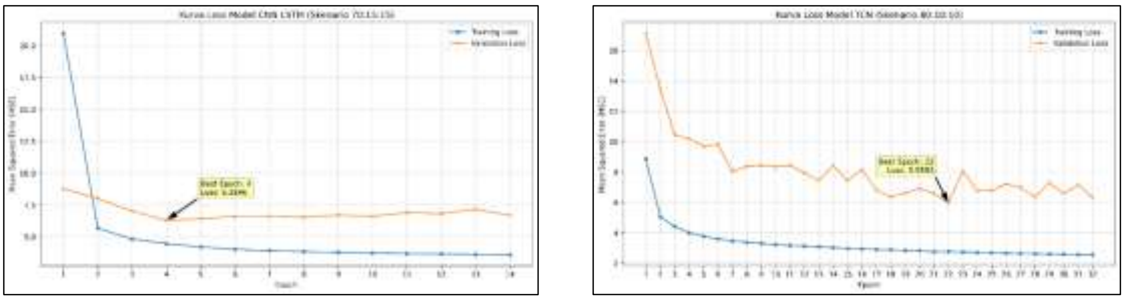


Figure 4. Pearson correlations among the seven predictors and MRT

3.2. TRAINING BEHAVIOUR AND AGGREGATE FORECAST ACCURACY

The 80:10:10 split yielded the most favorable error values for both architectures. CNN-LSTM reached its minimum validation loss of 6.1092 at epoch 10 and stopped after 20 epochs. TCN required 22 epochs to reach its best validation loss of 5.9592 and stopped after 32 epochs. The two best configurations therefore converged without persistent validation-loss growth, although the TCN curve fluctuated more during training because its deeper causal-convolutional structure contained substantially more parameters.



(a) (b)

Figure 5. Training and validation losses for the bess (a) CNN-LSTM and (b) TCN configurations

When the training share was reduced from 80% to 70%, CNN-LSTM performance deteriorated gradually, whereas TCN error increased more sharply. Under the 70:10:20 and 70:15:15 scenarios, TCN minimum validation losses exceeded 10, compared with values between 5.87 and 6.28 for CNN-LSTM. This suggests that the selected TCN configuration depended more strongly on the larger training set. A deeper dilated model can capture complex temporal interactions, but its additional capacity also increases its need for representative training examples.

Table 3. Test performance and residual bias across all model-split combinations

Split	Model	RMSE (°C)	MAE (°C)	MAPE	R^2	Mean residual (°C)
80:10:10	CNN-LSTM	2.8345	1.8874	5.76%	0.9509	−0.7803
80:10:10	TCN	2.8104	1.9570	5.96%	0.9518	−0.4863
70:10:20	CNN-LSTM	2.9475	2.0034	6.20%	0.9533	−1.1597

Split	Model	RMSE (°C)	MAE (°C)	MAPE	R^2	Mean residual (°C)
70:10:20	TCN	3.3778	2.4835	8.22%	0.9387	−1.3450
70:15:15	CNN-LSTM	3.0384	2.0512	6.32%	0.9484	−1.2339
70:15:15	TCN	3.4503	2.5110	8.19%	0.9334	−1.3973

All six experiments obtained R^2 values above 0.93, indicating that both architectures reproduced most of the temporal variance in gridded MRT. The highest R^2, 0.9533, occurred for CNN-LSTM under 70:10:20, even though that configuration did not have the lowest absolute error. This illustrates why R^2 should not be interpreted alone: it quantifies explained variation, whereas RMSE and MAE quantify temperature error.

Within the 80:10:10 split, TCN achieved the lowest RMSE (2.8104 °C) and the highest R^2 (0.9518). CNN-LSTM produced slightly lower MAE (1.8874 °C) and MAPE (5.76%). The RMSE difference between the two models was only 0.0241 °C, while the MAE difference was 0.0696 °C. Thus, their best aggregate performance was practically close, but the metrics reveal different error characteristics. TCN reduced the influence of larger errors, while CNN-LSTM slightly reduced the typical absolute error.

The CNN-LSTM results were more robust to the data-allocation change. Its RMSE increased from 2.8345 °C to 3.0384 °C between the most and least favorable scenarios, a difference of 0.2039 °C. TCN RMSE increased from 2.8104 °C to 3.4503 °C, a difference of 0.6399 °C. The recurrent memory mechanism combined with an initial convolutional extractor may have provided a more parameter-efficient representation for the available two-year dataset. Conversely, the selected TCN architecture was strongest when a larger portion of the chronology was available for training.

Table 4. Comparison of the two best configurations

Criterion	CNN-LSTM 80:10:10	TCN 80:10:10	Preferred result
RMSE	2.8345 °C	2.8104 °C	TCN
MAE	1.8874 °C	1.9570 °C	CNN-LSTM
MAPE	5.76%	5.96%	CNN-LSTM
R^2	0.9509	0.9518	TCN
Best validation loss	6.1092 at epoch 10	5.9592 at epoch 22	TCN
Mean residual	−0.7803 °C	−0.4863 °C	TCN
Overall interpretation	Lower average absolute and percentage errors; more stable across splits	Lower RMSE, higher R^2, lower validation loss, and smaller bias	TCN selected overall

3.3. RESIDUAL AND SPATIAL ERROR ANALYSIS

Residual distributions for both models were concentrated near zero but had negative means in every scenario. Because residual was defined as actual minus predicted MRT, these negative values indicate a modest average overprediction. Under the best split, TCN bias (−0.4863 °C) was closer to zero than CNN-LSTM bias (−0.7803 °C). The magnitude of bias increased when the test period became longer or the validation share increased, consistent with the deterioration in RMSE and MAE.

The actual-versus-predicted plots were concentrated near the 1:1 line, and hourly averages followed the observed daily cycle. Larger deviations occurred around rapid transitions and some extreme MRT values. These errors are physically plausible challenges because MRT responds rapidly to changes in direct radiation, cloud cover, and the transition between shaded and sunlit conditions. A 24-hour input window supplies the previous daily cycle, but ERA5 grid variables do not explicitly encode local shade, canyon geometry, or abrupt micro-scale cloud transitions.

Grid-level evaluation showed that error was not spatially uniform. The best TCN configuration produced lower RMSE across several central grid points, while errors increased toward some outer locations. Bias also followed a north–south gradient, with more negative values at some northern grids and values closer to zero or slightly positive toward the south. The pattern may reflect geographic differences in the radiation and atmospheric fields contained in the reanalysis, but it should not be interpreted as a direct urban-morphology effect because latitude, longitude, land cover, building geometry, and vegetation were not model inputs

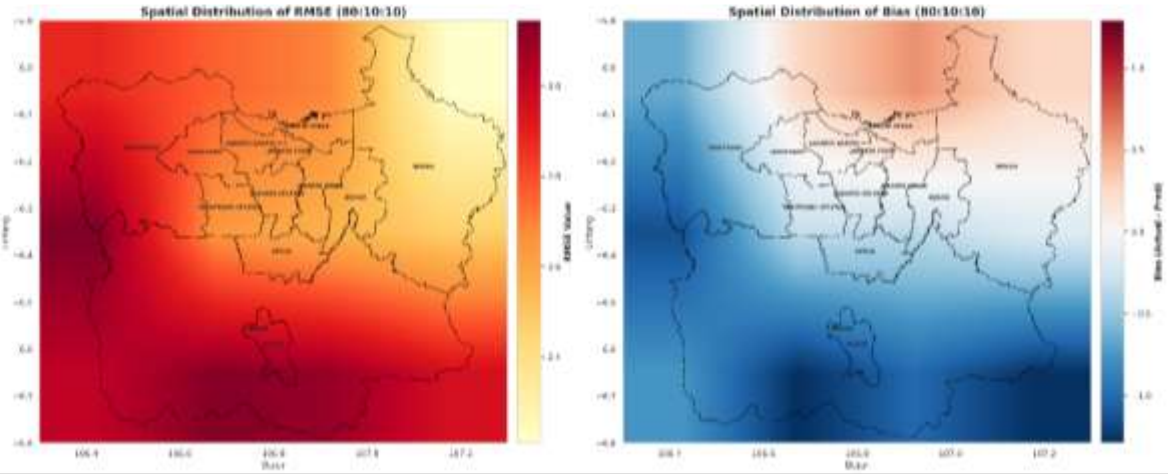


Figure 6. Grid-level RMSE and residual bias for the selected TCN 80:10:10 model

3.4. DISCUSSION, IMPLICATIONS AND LIMITATIONS

The findings demonstrate that multivariate deep-learning models can forecast the next-hour ERA5-HEAT MRT with high explained variance using only the previous 24 hours of radiation, temperature, and dew-point information. The strong performance is consistent with the physical composition of MRT: shortwave radiation variables dominated the linear association, while thermal variables supplied slower atmospheric context. The sequence models learned this recurring diurnal structure and its nonlinear deviations.

The comparison also confirms that architectural superiority is conditional rather than absolute. TCN was the strongest overall model under the most data-rich split, which agrees with the capacity of dilated causal convolutions to combine local and longer temporal relationships without recurrence (Bai et al., 2018; Lara-Benítez et al., 2020). However, CNN-LSTM retained better performance under reduced training proportions. Its convolutional layer compressed local interactions before the LSTM processed temporal dependencies, and this more compact architecture may have been less sensitive to the reduction in training samples.

The selected TCN offers a potential basis for a virtual MRT sensor or short-term thermal-monitoring component. A one-hour forecast could support dashboards for urban heat awareness, scheduling of outdoor work, and identification of periods requiring additional observation. Nevertheless, operational interpretation must remain cautious. The model predicts the ERA5-HEAT target at 0.25° resolution, not the radiant environment of a specific street or pedestrian. Direct deployment without local calibration could imply a level of spatial detail that the input data do not contain.

Several limitations affect generalization. First, the two-year period may not represent broader interannual variability or rare extreme events. Second, only one-step forecasting was evaluated. Third, hyperparameters were fixed rather than optimized through a nested time-series procedure. Fourth, computational efficiency was not compared, even though TCN contained nearly four times as many parameters as CNN-LSTM. Fifth, spatial coordinates and detailed urban characteristics were excluded, so the grid-error pattern cannot be fully explained.

The target itself introduces another limitation. ERA5-HEAT is a physically derived reanalysis product, and the models learn statistical relationships within that product family. High test accuracy therefore demonstrates successful emulation and short-term prediction of gridded MRT but does not establish equivalence with field-measured MRT. A field campaign using calibrated globe or directional-radiation instruments is needed to quantify local bias and to assess whether downscaling or transfer learning is required.

Despite these constraints, the experiment provides a controlled baseline for metropolitan tropical MRT forecasting. It shows the importance of chronological evaluation, reports architecture sensitivity to split composition, and combines conventional accuracy metrics with residual and spatial diagnostics. The results also indicate that model selection should consider the intended objective: CNN-LSTM is attractive when stable performance under limited training data is prioritized, while TCN is preferable when the lowest RMSE and smallest aggregate bias under a larger training set are the main criteria.

4. CONCLUTION

This study compared CNN-LSTM and TCN models for one-hour-ahead MRT forecasting across 12 Jabodetabek grid points using 210,528 hourly ERA5 and ERA5-HEAT records from 2023–2024. Seven radiation and near-surface atmospheric predictors were organized into independent 24-hour sequences and evaluated through three chronological train-validation-test splits. Both models achieved R^2 above 0.93 in all experiments. The 80:10:10 split was the best configuration for each architecture. TCN obtained the lowest RMSE of 2.8104 °C, R^2 of 0.9518, validation loss of 5.9592, and residual bias of −0.4863 °C. CNN-LSTM obtained the lowest MAE of 1.8874 °C and MAPE of 5.76% and degraded less when the training proportion was reduced. Based on RMSE, explained variance, validation loss, and bias, TCN 80:10:10 was selected as the overall best configuration. The results support the feasibility of deep-learning-based short-term MRT monitoring at reanalysis-grid scale. They do not yet justify direct pedestrian-level operational use because field validation, longer temporal coverage, and representation of urban morphology remain necessary.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the Indonesian Agency for Meteorology, Climatology, and Geophysics and the Master of Informatics Engineering Program at Universitas Pamulang for academic support. The author also thanks the Climate Early Warning team for technical discussions during data preparation and interpretation.

REFERENCES

Aviv, D., Guo, H., Middel, A., & Meggers, F. (2021). Evaluating radiant heat in an outdoor urban environment: Resolving spatial and temporal variations with two sensing platforms and data-driven simulation. *Urban Climate*, *35*, 100745. https://doi.org/10.1016/j.uclim.2020.100745

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*. https://doi.org/10.48550/arXiv.1803.01271

Briegel, F., Makansi, O., Brox, T., Matzarakis, A., & Christen, A. (2023). Modelling long-term thermal comfort conditions in urban environments using a deep convolutional encoder-decoder as a computational shortcut. *Urban Climate*, *47*, 101359. https://doi.org/10.1016/j.uclim.2022.101359

Brimicombe, C., Lo, C. H. B., Pappenberger, F., Di Napoli, C., Maciel, P., Quintino, T., Cornforth, R., & Cloke, H. L. (2023). Wet bulb globe temperature: Indicating extreme heat risk on a global grid. *GeoHealth*, *7*(2), e2022GH000701. https://doi.org/10.1029/2022GH000701

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Di Napoli, C., Barnard, C., Prudhomme, C., Cloke, H. L., & Pappenberger, F. (2021). ERA5-HEAT: A global gridded historical dataset of human thermal comfort indices from climate reanalysis. *Geoscience Data Journal*, *8*(1), 2–10. https://doi.org/10.1002/gdj3.102

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., . . . Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing, 24*(21), 16453–16482. https://doi.org/10.1007/s00500-020-04954-0

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Mining and Knowledge Discovery, 37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Lara-Benítez, P., Carranza-García, M., Luna-Romera, J. M., & Riquelme, J. C. (2020). Temporal convolutional networks applied to energy-related time series forecasting. *Applied Sciences, 10*(7), 2322. https://doi.org/10.3390/app10072322

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lindberg, F., Holmer, B., & Thorsson, S. (2008). SOLWEIG 1.0—Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings. *International Journal of Biometeorology, 52*(7), 697–713. https://doi.org/10.1007/s00484-008-0162-7

Matzarakis, A., Rutz, F., & Mayer, H. (2010). Modelling radiation fluxes in simple and complex environments: Basics of the RayMan model. *International Journal of Biometeorology, 54*(2), 131–139. https://doi.org/10.1007/s00484-009-0261-0

Oke, T. R., Mills, G., Christen, A., & Voogt, J. A. (2017). *Urban climates*. Cambridge University Press. https://doi.org/10.1017/9781139016476

Ren, J., Shi, K., Li, Z., Kong, X., & Zhou, H. (2023). A review on the impacts of urban heat islands on outdoor thermal comfort. *Buildings, 13*(6), 1368. https://doi.org/10.3390/buildings13061368

Thorsson, S., Lindberg, F., Eliasson, I., & Holmer, B. (2007). Different methods for estimating the mean radiant temperature in an outdoor setting. *International Journal of Climatology, 27(*14), 1983–1993. https://doi.org/10.1002/joc.1537

Zhou, S., Guo, S., Du, B., Huang, S., & Guo, J. (2022). A hybrid framework for multivariate time series forecasting of daily urban water demand using attention-based convolutional neural network and long short-term memory network. *Sustainability, 14*(17), 11086. https://doi.org/10.3390/su141711086