26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Impute-Net: A Deep Learning Approach to Missing Observations in Satellite Imagery

Achmad Raflie Pahlevi [a], **Akbar Ghobakhlou** [a], **Jacqueline Whalley** [b]

[a]*Auckland University of Technology, NZ*
[b]*Lincoln University, NZ*
Email: achmad.pahlevi@autuni.ac.nz, akbar.ghobakhlou@aut.ac.nz

Abstract: The impact of natural disasters, including damage to infrastructure and loss of human life, underscores the importance of accurate predictions of extreme weather. Unreliable datasets, such as those with missing observations, hinder the prediction process, preventing machine learning and deep learning from producing notable weather predictions. Most contemporary approaches to prediction use data from remote sensing and in particular satellite imagery. Thus, development of precise methods to impute missing values in satellite imagery is essential to weather prediction. To the best of our knowledge, there is no deep learning method that can generate complete missing frames. Most deep learning studies only handle inpainting or filling in the missing pixels or small patches of the frames. This paper introduces a deep learning method called Impute-Net which is inspired by the U-Net architecture. Impute-Net uses both pre- and post-gap information on two different encoding branches to generate missing frames. Experimental results demonstrate that Impute-Net, when compared with UNet, achieves superior imputation for shorter temporal gaps ($x \leq 3$), yielding the lowest error metrics (RMSE and MAE) and the highest structural fidelity (PSNR and SSIM). For longer gaps ($x \geq 5$), the original U-Net model exhibits lower error values due to its smoothing effect; however, this comes at the cost of reduced structural accuracy. The high-quality and precise imputations produced by Impute-Net suggest its potential to significantly improve the accuracy of weather prediction models.

Keywords: *Impute-Net, U-Net, Deep Learning, Imputation, Satellite Imagery*

1 INTRODUCTION

Satellite imagery can address the issue of spatially sparse and limited surface observations and has been used by public officials in monitoring disasters such as hurricanes, floods, and wildfires (Akhyar et al., 2024). Despite the coverage of satellite imagery, some missing values can hinder the accuracy and dependability of analysis and modelling studies (Kotan and Kırışoğlu, 2025; Hu et al., 2024). In weather forecasting, unlike physics-based models that can tolerate some missing data, deep learning (DL) and machine learning (ML) models require complete temporal sequences. Without proper preprocessing strategies, such as handling missing values, ML and DL cannot achieve optimal results (Arefin and Masum, 2024).

Approaches to handling missing values through imputation include using traditional statistical methods (e.g. regression), ML and DL. Traditional statistical methods are constrained by their reliance on strong distributional assumptions, which are often violated in geospatial data (Antariksa et al., 2023). Statistical imputation methods also often perform poorly in terms of model accuracy and stability, particularly with data such as weather data, which has high dimensionality and complex features. Consequently, this can affect model accuracy and stability (Qin et al., 2024).

ML and DL are increasingly used for imputation because they capture complex patterns in large datasets, including spatial patterns of satellite images. A Convolutional Neural Network (CNN) can be used to transform raw data into a spatiotemporal image and reconstruct the complete picture by imputing missing frames (Zhuang et al., 2019). A CNN contains a convolutional layer for feature extraction, a pooling layer to decrease spatial dimensions, an activation function to produce the final output, and a fully connected layer that integrates high-level features nonlinearly (Bhatt et al., 2021).

The advanced CNN architecture known as U-Net is increasingly popular for imputing missing frames because it offers accurate segmentation and image extrapolation while needing fewer training images (Zhao et al., 2020). Firdaus et al. (2024) employed U-Net to impute missing values by training the network on complete datasets, learning to encode the input data into a lower-dimensional space and subsequently decoding it back into the original data space, allowing it to reconstruct any incomplete data. The other widely used deep learning model, Transformer, excels in handling high-dimensional data with complex temporal dependencies and delivers strong performance for dynamic signals, where traditional methods often fall short (Jungo et al., 2024). In observing environmental conditions, Cesar et al. (2023) used a Transformer to maintain the data quality of solar irradiance despite sensor failure.

Most recent DL work on missing data focuses on inpainting pixels rather than reconstructing entire frames. Pondaven et al. (2022) used Convolutional Neural Processes (ConvNP), which employs a latent variable to learn details from a contextual dataset to paint in the missing areas of an image using the available data pixels. Another example of inpainting is the work of Gong et al. (2023) who employed a dilated and self-attention U-Net (DSA-UNet), incorporating a self-attention block into the first two upscaling layers to learn patterns for restoring missing pixels in weather radar images.

This study proposes Impute-Net, a U-Net-inspired architecture that reconstructs entire missing frames using pre- and post-gap context. The approach preserves spatiotemporal continuity, enabling DL models to learn from complete sequences for reliable nowcasting. Details are provided in Section 2.

2 METHODOLOGY

2.1 Data Collection and Preparation

The Impute-Net model will be built using Himawari-9 Satellite imagery from the Japan Meteorological Agency (JMA). This satellite enables ongoing monitoring of atmospheric events, including tropical cyclones, thunderstorms, heavy rainfall, and significant snowfall, across East Asia and the Western Pacific (Bessho et al., 2016). For this study, we compiled data from 1 to 31 July 2025 with a spatial resolution of 5 km and a temporal resolution of 10 minutes, yielding 144 images per channel per day or 2,304 images/day across 16 channels. In total, the dataset comprises 71,424 images (70,192 valid and 1,232 null) used for training and evaluation.

Data preparation, for this research, is divided into two steps: normalisation and data partitioning. The data is split into train, validation, and test subsets to prevent data leakage, with the proportions of 70%, 10%, and 20% respectively. During training, only the training subset is used to update the model's weights. In contrast, the validation subset is used at the end of each epoch to monitor performance and to prevent overfitting. Finally, the testing subset remains completely untouched throughout the training process and is used only in the evaluation phase to provide an unbiased assessment of the model's generalization ability.

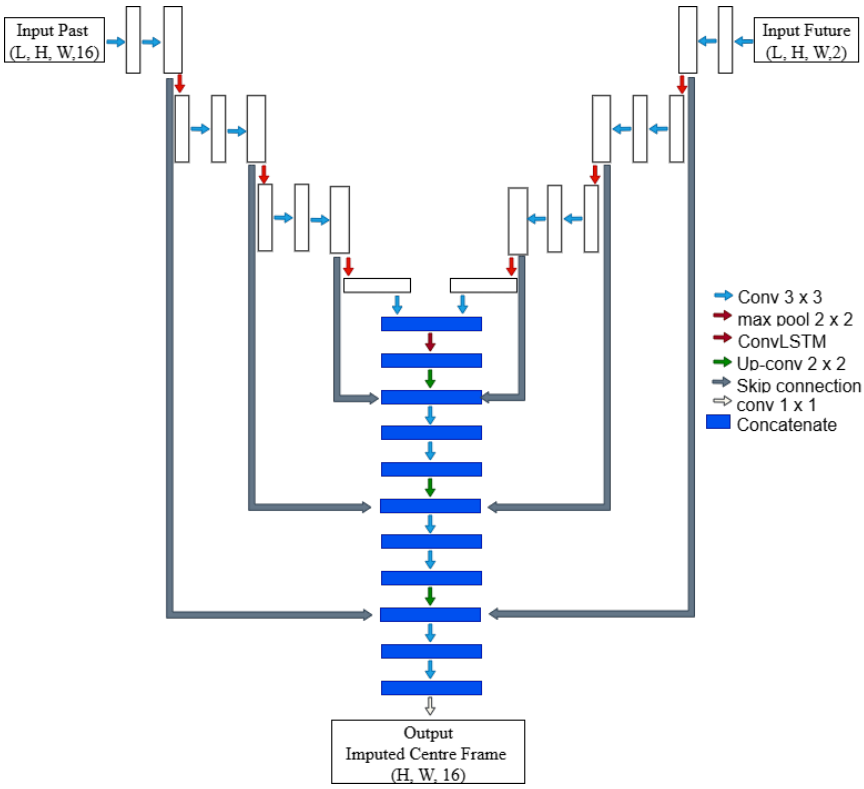

Figure 1. Model Architecture of Impute-Net. Input past/future $(7, 281, 281, 17)$**, where** $7 = 3$ **past** $+1$ **masked centre** $+3$ **future frames. Bottleneck ConvLSTM** $(7, 36, 36, 128) \rightarrow$ **centre slice** $(36, 36, 128) \rightarrow$ **output** $(281, 281, 16)$

2.2 Model Architecture

Figure 1 shows the architecture of our proposed novel model. We employ two sources as input and treat them as two distinct branches rather than two separate channels. These two sources represent the images from both the immediate past and the future of the missing observation, with an equal number of past $(P = Z_{(t-k)}, \ldots, Z_{(t-1)})$ and future $(F = Z_{(t+1)}, \ldots, Z_{(t+k)})$ to input the missing frame at t (Z_t). The spatial dimensions for these two input sources, as well as the output of this model, are denoted as H and W.

Three encoding levels are used to capture spatial structures and abstract features into meaningful patterns. Processing past and future inputs independently preserve their unique characteristics, with each branch denoted as H:

$$H_l^p = \text{Enc}_l(P), \quad H_l^f = \text{Enc}_l(F), \quad l = 1, \ldots, L \tag{1}$$

The bottleneck integrates high-level features from both branches, where a ConvLSTM capture the temporal dependencies by concatenating the past encoding $H_{L_e}^p$ and the future encoding $H_{L_e}^f$, denoted by the concatenation operator $||$:

$$B = \text{ConvLSTM}(H_{L_e}^p \parallel H_{L_e}^f) \tag{2}$$

The decoder upsamples the concatenated features back to the original size. Upsampling (Up) restores spatial scale, while skip connections (S) recover details lost during downsampling . The decoding (D) for each layer is notated as:

$$D_{l-1} = \Phi_l \left(\text{Up}(D_l) \parallel S_{l-1}\right), \quad D_l = B \text{ for } l = L_e \tag{3}$$

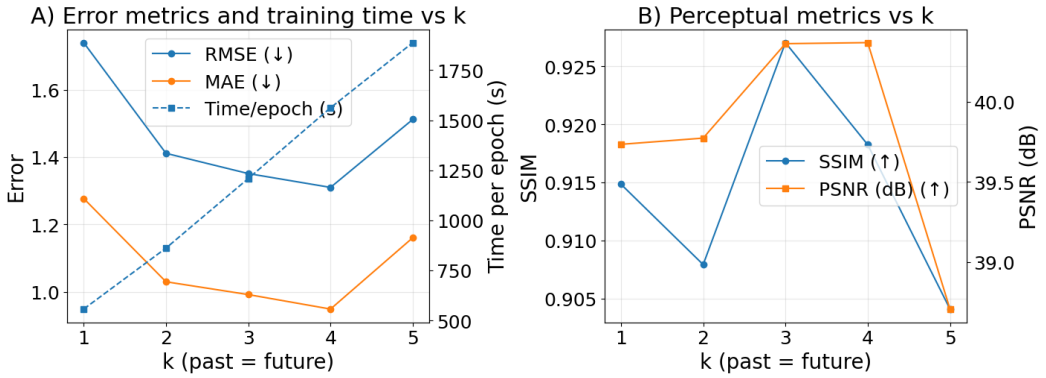

Figure 2. The comparison of Model Performance with the Number of K Window

The final convolutional layer will be used to return the upsampling output to the original size $H \times W$, with only one channel, and it is notated as:

$$\hat{Y} = \text{Conv}_{1\times1}(D_0) \tag{4}$$

2.3 Training Procedure and Evaluation

Each training sample is generated by masking the centre frame at time t within a window of k past and future steps. The masked frame is replaced with NaN or null and flagged, ensuring the model never sees the masked frame's content. The model reconstructs the missing frame from the temporal context based on the pre- and post-gaps, while the ground-truth centre frame $Y = \tilde{X}_{t_0}^{\text{true}}$ serves as the target. The loss between the prediction and ground truth is then used to update model weights. For this research, we used spatial resolution of 281×281 pixels ($\approx 7.9 \times 10^4$ pixels per frame at 5 km scales) across 16 channels plus one availability flag. The Impute-Net model contains 2.99 million parameters (11.41 MB). Training was conducted on an NVIDIA L4 GPU with 24 GB VRAM, where the peak for a batch size of four was approximately 17.6 GB.

To evaluate the performance of the model, it will be compared with the U-Net and the Transformer. The 3D U-Net processes stacked previous frames without future input or using ConvLSTM at its bottleneck. The Transformer baseline is implemented as a spatiotemporal encoder-decoder with self-attention layers across previous frames. In contrast, Impute-Net integrates both past and future frames through a ConvLST bottleneck. Both models are assessed on their ability to impute one to six consecutive missing frames, consistent with prior studies in filling missing Sea Surface Temperature (SST) pixels using U-Net (Landt-Hayen et al., 2023) and missing solar irradiance data using Transformer (Cesar et al., 2023).

Five different metrics will be employed to evaluate the performance of these models, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). RMSE, MAE, and PSNR will be calculated and compared pixel by pixel for the predicted frames, while SSIM will be calculated and compared for the overall structure of the expected frame.

3 RESULT AND DISCUSSION

3.1 Model Development

The following experiment is conducted to find the best window gap combination for the prior and future branches as input. The results of experimenting with the model using the first five epochs with different window sizes (k) are shown in Figures 2. These results are averaged across all 16 channels. Figure 2 shows that increasing k from 1 to 4 reduces the error in both RMSE and MAE, but the error rises again beyond k = 5. Furthermore, the cost of time per epoch increases linearly with the number of k, indicating that using more windows affects the effectiveness of the training process. Figure 2.b shows that k = 3 is the window with the highest SSIM and also has a high PSNR, whereas k = 4 has a lower SSIM when compared to the window of 3.

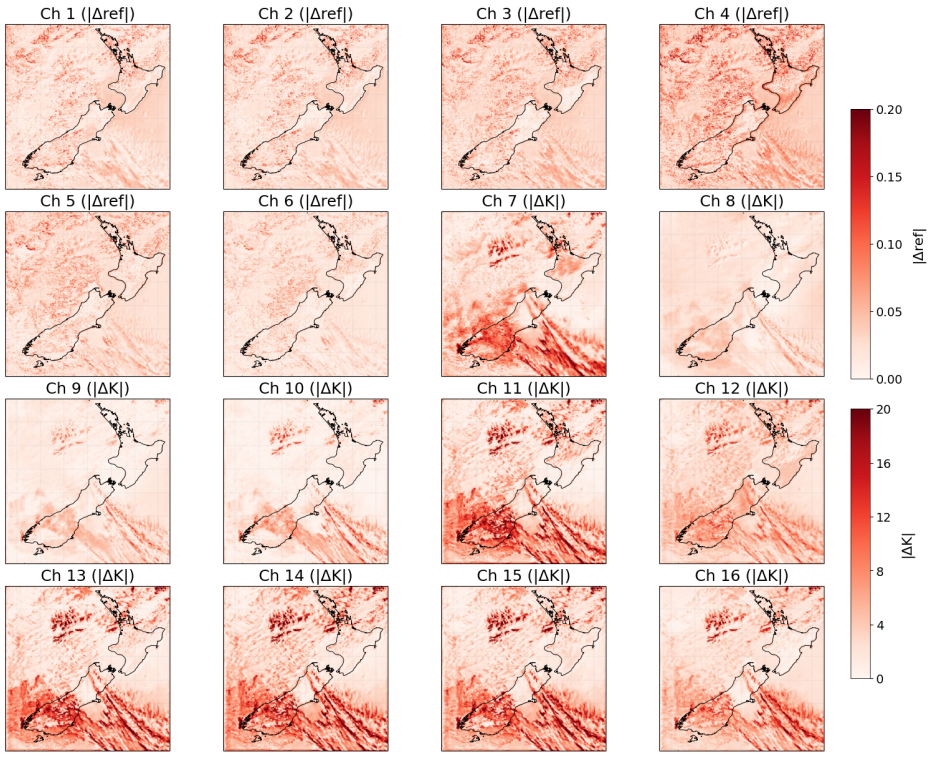

Figure 3. Absolute error of Image Reconstruction using Impute-Net across all 16 Himawari-9 channels for the case of a single missing frame ($k_{missing} = 1$) with context window size $k = 3$

3.2 Image Reconstruction and Evaluation

The performance of imputation using Impute-Net is illustrated in Figure 3, which compares the difference between the prediction and the ground truth. Visible bands (Channels 1-6), shown in the first six images from the top, indicate the magnitude of reflectance error. These results demonstrate that Impute-Net precisely provides imputation with generally low error over clear skies and land, while incurring a slight error of less than 0.20 for areas with clouds.

The brightness temperature band is shown from Channel 7 to 16 in the bottom two rows, and the other two images from the second row, illustrating the differences in BT around 0 to 20 K. For the upper-tropospheric water vapour band (Channels 8-10), the Impute-Net model exhibits a slight error of approximately 8 K in the southeast of New Zealand, highlighting the challenge in capturing jet-level advection and vertical moisture structure. Furthermore, the Short-Wave IR (Channel 7) and the Window Bands (Channels 13-16) exhibit errors in the same cloud systems, specifically in small clouds or at the edges of convective clouds.

In image reconstruction, the error increases at the edge of the cloud and in areas with more intense clouds. Furthermore, using a single training to determine the weight of the Impute-Net model results in errors that vary across different channels; some channels experience low error rates, while others have higher error rates. However, Impute-Net exhibits lower error in areas with clear sky or land, indicating that the model can effectively distinguish between regions with and without clouds. Overall, Impute-Net can impute and reconstruct the missing frames of the Himawari Satellite, achieving high structural fidelity with sharp edges and avoiding the smoothing mean error.

Figure 4 illustrates four different metrics for evaluating the performance of Impute-Net (blue) compared to other DL models in imputing one to six consecutive missing frames. The results are averaged across all 16 Himawari channels. The two top images, SSIM and PSNR, reflect the weakening temporal correlations as the holes widen. Even so, Impute-Net successfully achieves the highest SSIM (0.86–0.94) and the highest PSNR (36.8–42.8) compared to other models. The SSIM differences between Impute-Net and other models increase (from 0.02 to 0.06), while the PSNR decreases (from 1.8 to 0.8) as the gaps widen. It illustrates that

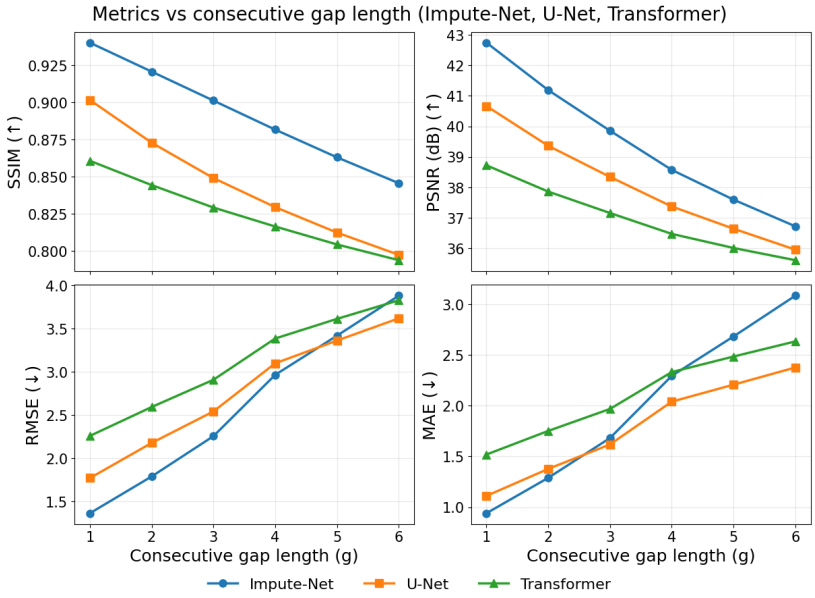

Figure 4. The comparison of Different Models for Imputing Different Numbers of Consecutive Gaps

the imputation using Impute-Net shows sharper cloud edges and better texture recovery.

For short to moderate gaps (≤ 3), Impute-Net demonstrated the lowest RMSE/MAE compared to the other models. For longer gaps (≥ 3), the U-Net become slightly better with lower error for both RMSE and MAE, while still trailing in SSIM and PSNR. Meanwhile, Transformer shows the lowest performance compared to the other two models in all metrics. This result demonstrates that the U-Net produces smoother reconstructions closer to the mean for a larger gap, but is less accurate structurally.

Impute-Net requires pre- and post-gap frames as input. While this bidirectional conditioning yields the highest SSIM/PSNR for short gaps, its performance becomes less robust as the consecutive gap length increases. When dealing with longer gaps, the post-gap frames become unavailable. Without the post-gap context, the post-gap branches contribute less signal and more noise, causing a mismatch between training and testing. This leads to small misplacements (higher errors) while the edges remain sharp (good SSIM). For operational meteorology, where the structure matters more than the minor error, Impute-Net outperforms the U-Net, even for the longer gap.

The trade-off between structure and smoothness is clear in Impute-Net's performance. As the gaps widen, Impute-Net maintains coherent morphology effectively, though this leads to larger pixel errors due to sharper predictions when minor misalignments occur over longer gaps. In operational meteorology, structural fidelity, such as with cloud edges and convective cores, is more critical than pixel-by-pixel accuracy, thus favouring Impute-Net over U-Net.

4 CONCLUSION AND FUTURE WORK

The results obtained in this research highlight the ability of Impute-Net to impute missing frames of satellite observations, particularly those from the Himawari Satellite. Compared to other models, Impute-Net exhibits the highest structural fidelity, achieving high PSNR and SSIM values for all gaps. Impute-Net is observed to have the lowest errors (both RMSE and MAE) for smaller gaps, and slightly higher for larger gaps, compared to U-Net. U-Net can outperform Impute-Net in terms of error, while still achieving lower SSIM/PSNR, indicating a smoother, mean-regressing reconstruction.

The results reveal two main limitations: first, a higher error rate than other models as the number of consecutive gaps increases; second, varying absolute errors across different channels. To address these challenges, we propose some approaches for future research. To address long consecutive gaps, we suggest using iterative imputation with curriculum learning to impute the missing frame when more than one successive frame is missing. The iterative imputation would impute the missing frame starting from the closest gap to the available

data, either pre- or post-gap, and iteratively move to the centre of the gap.

To address the limitation of high error variability across different channels in the Impute-Net product, we propose channel-wise weighting for improved generalisation and physical fidelity. By employing uncertainty-weighting, we treat each channel individually, so that the channel with higher noise (or uncertainty label) is down-weighted without being ignored. This method will restructure the training process to help reduce cross-channel variance in error.

ACKNOWLEDGMENT

This research is conducted under the Manaaki New Zealand Scholarship, an award funded by the New Zealand Government to enlighten academic and professional development through international education. The Himawari satellite image dataset is supplied by JAXA JMA.

REFERENCES

Akhyar Akhyar, Asyraf Zulkifley Mohd, Lee Jaesung, Song Taekyung, Han Jaeho, Cho Chanhee, Hyun Seunghyun, Son Youngdoo and Hong Byung-Woo (2024), Deep artificial intelligence applications for natural disaster management systems: A methodological review, *Ecological Indicators* **163**, 112067.

Antariksa Gian, Muammar Radhi, Nugraha Agung and Lee Jihwan (2023), Deep sequence model-based approach to well log data imputation and petrophysical analysis: A case study on the west natuna basin, indonesia, *Journal of Applied Geophysics* **218**, 105213.

Arefin Muhammed Nazmul and Masum Abdul Kadar Muhammad (2024), A probabilistic approach for missing data imputation, *Complexity* **2024**(1), 4737963.
URL: *https://doi.org/10.1155/2024/4737963*

Bessho Kotaro, Date Kenji, Hayashi Masahiro, Ikeda Akio, Imai Takahito, Inoue Hidekazu, Kumagai Yukihiro, Miyakawa Takuya, Murata Hidehiko, Ohno Tomoo, Okuyama Arata, Oyama Ryo, Sasaki Yukio, Shimazu Yoshio, Shimoji Kazuki, Sumida Yasuhiko, Suzuki Masuo, Taniguchi Hidetaka, Tsuchiyama Hiroaki, Uesawa Daisaku, Yokota Hironobu and Yoshida Ryo (2016), An introduction to himawari-8/9 - japan's new-generation geostationary meteorological satellites, *Journal of the Meteorological Society of Japan. Ser. II* **94**(2), 151–183.

Bhatt Dulari, Patel Chirag, Talsania Hardik, Patel Jigar, Vaghela Rasmika, Pandya Sharnil, Modi Kirit and Ghayvat Hemant (2021), Cnn variants for computer vision: History, architecture, application, challenges and future scope, *Electronics* **10**(20), 2470.

Cesar Llinet B., Manso-Callejo Miguel-Ángel and Cira Calimanut-Ionut (2023), Bert (bidirectional encoder representations from transformers) for missing data imputation in solar irradiance time series.

Firdaus Firdaus, Nurmaini Siti, Tutuko Bambang, Rachmatullah Muhammad Naufal, Islami Anggun, Darmawahyuni Annisa, Sapitri Ade Iriani, Ais'sy Widya Rohadatul, Karim Muhammad Irfan, Fachrurrozi Muhammad and Zarkasi Ahmad (2024), Xu-neti: Simple u-shaped encoder-decoder network for accurate imputation of multivariate missing data, *Franklin Open* **8**.

Gong Aofan, Chen Haonan and Ni Guangheng (2023), Improving the completion of weather radar missing data with deep learning, *Remote Sensing* **15**(18), 4568.

Hu Hanwen, Qian Shiyou, Yang Dingyu, Cao Jian and Xue Guangtao (2024), Iterative time series imputation by maintaining dependency consistency, *ACM Trans. Knowl. Discov. Data* **19**(1), Article 6.

Jungo Janosch, Xiang Yutong, Gashi Shkurta and Holz Christian (2024), Representation learning for wearable-based applications in the case of missing data, *arXiv preprint arXiv:2401.05437* .

Kotan Kurban and Kırışoğlu Serdar (2025), Cyclical hybrid imputation technique for missing values in data sets, *Scientific Reports* **15**(1), 6543.

Landt-Hayen Marco, Kröger Peer, Rath Willi and Claus Martin (2023), Reconstruct geospatial data from ultra sparse inputs to predict climate events.

Pondaven Alexander, Bakler Märt, Guo Donghu, Hashim Hamzah, Ignatov Martin and Zhu Harrison (2022), Convolutional neural processes for inpainting satellite images, *arXiv preprint arXiv:2205.12407* .

Qin Xiwen, Shi Hongyu, Dong Xiaogang, Zhang Siqi and Yuan Liping (2024), Improved generative adversarial imputation networks for missing data, *Applied Intelligence* **54**(21), 11068–11082.

Zhao Y., Shangguan Z., Fan W., Cao Z. and Wang J. (2020), U-net for satellite image segmentation: Improving the weather forecasting, *in* '2020 5th International Conference on Universal Village (UV)', pp. 1–6.

Zhuang Yifan, Ke Ruimin and Wang Yinhai (2019), Innovative method for traffic data imputation based on convolutional neural network, *IET Intelligent Transport Systems* **13**(4), 605–613.

