Response to Reviewers

Manuscript Title: Precipitation Nowcasting Based on Convolutional LSTM with Spatio-Temporal Information Transformation Using Multi-Meteorological Factors

Dear Reviewer,

We would like to express our sincere gratitude to the Editor and the Reviewers for dedicating their time to evaluating our manuscript. We deeply appreciate the Reviewer's encouraging summary, acknowledging the relevance of our multimodal SEVIR setup and the value of our ablations. The constructive and meticulous feedback provided has been immensely helpful in improving the rigor, clarity, and overall quality of our paper.

We have carefully considered all the comments and revised the manuscript accordingly. Below, we provide a point-by-point response to each of the Reviewer’s concerns.

Part I: Response to Major Concerns

Comment 1: *The manuscript invokes delay embedding, STI equations, and conjugate duality... Please reframe the method as STI-inspired/STI-motivated dual learning and clearly separate the mathematical motivation from the learned implementation.*

Response: We sincerely thank the reviewer for this mathematically rigorous observation. We completely agree that our neural network acts as a data-driven approximator rather than an explicit numerical solver for the STI equations. Following your highly constructive suggestion, we have reframed our methodology throughout the manuscript. We now accurately describe our model as an "STI-inspired dual learning framework" that "implicitly approximates spatiotemporal mappings," ensuring a clear boundary between the mathematical motivation and the neural implementation.

To clarify our original thought process: because deep learning fundamentally operates as an implicit function-fitting mechanism, we utilized the model to approximate the STI mapping function itself. We originally viewed this approximation process as implicitly "solving" the conjugate STI function. However, we concede that the term "solve" can be mathematically misleading in this context, and your suggested phrasing is much more accurate and appropriate for a deep learning audience.

Comment 2: *The forecast target is VIL, not direct surface precipitation... the manuscript should consistently describe the task as VIL nowcasting...*

Response: We appreciate the reviewer pointing out this critical distinction. We fully acknowledge that VIL is a radar-derived proxy for convective intensity and strictly differs from direct surface precipitation or rainfall rates. We have thoroughly revised the manuscript to consistently describe our specific task as "VIL nowcasting" and removed any misleading terminology regarding surface precipitation.

Furthermore, to address the underlying concern regarding the real-world applicability of our method to actual rainfall, we have proactively conducted an additional validation experiment. We trained our model on an independent, real-world precipitation dataset and compared it against traditional forecasting methods. The results demonstrate that our method maintains its effectiveness and outperforms traditional approaches in predicting true rainfall volumes. We have added this supplementary experiment to the manuscript and explicitly outlined the expansion to broader real-precipitation datasets as a key focus for our future work.

Comment 3: *The model appears deterministic, yet CRPS and BSS are reported. Please either... or remove/demote them from the main results.*

Response: We appreciate the reviewer's strict assessment of our evaluation metrics. You are entirely correct that our model generates deterministic point forecasts, making the interpretation of continuous probabilistic scores like CRPS and BSS mathematically unjustified in this context. Following your recommendation, we have completely removed the CRPS and BSS metrics from the tables and the main text to maintain the absolute rigor of our deterministic evaluation.

Comment 4: *Please add a baseline setup table. At minimum, it should specify input channels, preprocessing... If official implementations were used with minimal changes, state this explicitly.*

Response: Thank you for highlighting the need for transparency and reproducibility. We have added explicit statements in the manuscript to clarify our baseline setups. To guarantee absolute fairness, all baseline models were trained and evaluated using their exact official open-source implementations without any architectural or hyperparameter modifications. Furthermore, the input datasets (including channels and preprocessing) fed into the baselines were strictly identical to those used for our proposed method.

Because we adhered 100% to the official configurations for each respective baseline, their hyperparameters and specific training setups vary drastically by architecture. Compiling all these disparate, algorithm-specific parameters into a single coherent table proved impractical and potentially confusing. Instead, we have explicitly documented in the text that we strictly followed the parameters established in the official baseline repositories and corresponding papers. This also applies to the generative model DGMR, where the implementation is completely aligned with the official paper, changing only the input data stream to the SEVIR dataset.

Comment 5: *The random SEVIR event split may be standard, but it may still overestimate generalization... Confine all claims to the specific conditions tested... Report the count and percentage of no-rain events excluded...*

Response: 1. **Regarding Claims and Generalization:** We completely agree with your assessment. A random split might inadvertently capture temporal or storm-system correlations. Because the SEVIR dataset is structured around discrete weather events rather than continuous timelines, isolating specific seasons or geographical regions for a held-out test set poses significant structural challenges. We have now explicitly acknowledged this limitation in the manuscript. As suggested, we have carefully added qualifiers to our conclusions (specifying dataset, forecast horizons, and event intensities) to avoid overclaiming. Incorporating specific seasonal and regional hold-out evaluations has been formally added to our Future Work section.

\2. **Regarding the "No-Rain Events" :** We owe the reviewers a deep apology for a serious descriptive error in our original text. In fact, from the original pool of approximately 13,000 SEVIR events, we strictly filtered for events due to a lack of sensor data or a lack of a continuous four-hour time gap. The total number of events excluded was approximately 2% of the dataset, leaving us with 12,000 complete events, of which we selected 10,000 for experimentation. We describe this practice of ours in the text as "we excluded events without rainfall," and we correct this wording in Section 4.1 to accurately reflect our data quality control process and to confirm that the natural category balance of empty events was preserved.

Comment 6: *Claims should be qualified by lead time, metric, and target variable. Longer-lead results show that other models achieve better MSE/PSNR...*

Response: We fully agree with this insightful feedback. We have carefully revised our manuscript to remove any implications of uniform superiority. Specifically, in the Abstract, Introduction, and Conclusion, we have strictly bounded our claims, explicitly stating that our model primarily improves structural and intensity-focused metrics (CSI and HSS) for short-term forecasting. Additionally, we have added a more balanced discussion in Section 5.3, transparently highlighting the respective advantages of our method alongside the strengths of Transformer-based baselines (which inherently excel in MSE/PSNR over longer horizons).

Part II: Response to Minor Issues

Comment 1: *The introduction is long and verbose. Please reduce it where possible.*

Response: We have thoroughly revised and streamlined the Introduction, removing overly verbose meteorological background information and bringing the focus to the core contributions much earlier in the text.

Comment 2: *Define NWP, LSTM, SOTA, and other abbreviations at first use.*

Response: We have carefully proofread the manuscript and ensured that all abbreviations (including NWP, LSTM, and SOTA) are fully spelled out and defined upon their first occurrence.

Comment 3: *Add confidence intervals to the metrics in Tables 4 and 5...*

Response: We have updated our results to include confidence intervals for the evaluation metrics, utilizing the testing data from our test set to compute the variance, thereby providing a clearer picture of the statistical stability of our performance.

Comment 4: *Improve the readability of large multi-panel figures. Figure 3... requires a layer-by-layer table...*

Response: We entirely agree that a layer-by-layer breakdown is vital for reproducibility. We have added a comprehensive architectural parameter table to the manuscript detailing the channel counts, tensor shapes, and activation functions for each layer of the network.

Comment 5: *Section 4.3 reports training time and inference speed; add a compact comparison with key baselines...*

Response: We have added the requested computational efficiency data. We expanded our discussion in the text and included the training epochs and inference speeds (in frames per second/seconds per sequence) for all evaluated baseline models alongside our own, providing a clear operational comparison.

Part III: Response to Questions for Clarification

Q1. How exactly are CRPS and BSS computed from model outputs?

Response: As acknowledged in our response to Major Concern 3, interpreting these metrics was mathematically inappropriate for our deterministic point forecasts. We have completely removed CRPS and BSS from the manuscript.

Q2. Do all baselines use the same multimodal inputs and preprocessing?

Response: Yes. All baseline models utilized the exact same multimodal inputs and preprocessing steps as our proposed method.

Q3. Were all baselines retrained using the same split and validation criterion?

Response: Yes. All baselines were retrained from scratch using our exact same train/validation/test splits and identical validation criteria.

Q4. How many no-rain events were removed, as a count and percentage?

Response: We have revised the relevant statement in the text. As detailed in our response to Main Concern 5, the statement in the original text is a description error. We excluded about 300 events (about 2% of the original sample pool) strictly based on missing sensor data, rather than excluding weather events due to lack of rainfall.

Q5. Can the method be tested on held-out temporal or spatial subsets?

Response: Because the SEVIR dataset is event-based rather than a continuous chronological or mapped geospatial stream, creating strict temporal (seasonal) or spatial held-out subsets is not structurally supported without fundamentally altering the dataset. We have noted this as a limitation and earmarked it for future validation on continuous datasets.

Q6. What is the precise implementation of the alpha weighting in ADGLoss?

Response: We apologize if the practical implementation was not sufficiently clear in the text. While Equation (7) provides the mathematical definition, we have added a clarification in Section 3.3 detailing that, programmatically, the adaptive weight alpha is computed dynamically per batch during the forward pass. This allows the gradient to scale adaptively based on the specific precipitation intensity of the batch currently being processed.

We hope these revisions and explanations satisfactorily address your concerns. Thank you once again for your invaluable guidance in improving our work.

Sincerely,

The Authors