

1 Reply to referee 1

General Comments

This is a well-structured research paper on introducing fire breaks to simulations that predict fire spread, while detailing the different machine learning algorithms that take this important element into account. While the research is novel and the methodology is reproducible, the impact it may have in actual wildfire situations is questionable. The approach that the authors propose can certainly be used to better test/understand wildfire dynamics in virtual environments, and may even be used to train first responders, however it falls short of becoming a core application in case of a live wildfire mainly for two reasons: 1- Wildfire propagation is greatly related to wind, and with climate change we see higher uncertainty and extremes in wind patterns. Coarse resolution wind data may grossly underestimate what is going on in the actual wildfire scene, as not only wind speed, but wind gusts during an active fire are major players in fire severity and spread (starting new ignitions in forest patches that are far away from the main ignition zone, regardless of a fire break). Also, a severe wildfire can produce its own wind patterns, shift the current wind direction, or increase its intensity. The authors state that they are aware of this shortcoming in their study. But this undermines their claim that their methodology may prove as an effective strategy to reduce wildfire impacts, since wind can also affect how temporary breaks are (or can be) deployed. 2- Wildfire propagation is also greatly related to topography, but the authors mention that they did not have a comprehensive landscape dataset available, so the ConvLSTM simulations were run with null input for landscape.

As a learning and potential training tool, despite its limitations, I find the work insightful with room for improvement, a possible first step towards developing a global dynamic simulator that considers fire breaks while projecting potential wildfire spread and provide valuable insight for effective containment. The clear methodology helps the simulations' reproducibility, and aid researchers to test it in their study areas. However, speed vs accuracy between CA and DL models needs to be carefully considered, as one should not replace the other.

Thank you for the positive evaluation and for the detailed and constructive comments. We have carefully considered all points raised and revised the manuscript accordingly. A point-by-point response is provided below.

We agree that wind variability and complex topographic influences play a critical role in real-world wildfire propagation. In the current implementation, wind conditions are assumed to be spatially constant during each simulation in order to control the dimensionality of the training space. Introducing fully dynamic and high-resolution wind fields would require generating a substantially larger number of CA scenarios to adequately cover the expanded parameter space for training.

As clarified in the revised manuscript, the primary objective of this study is to provide a proof of concept demonstrating that a ConvLSTM-based surrogate model can successfully learn wildfire spread dynamics with dynamic firebreak deployment. The framework is flexible, and future extensions can incorporate variable wind directions, gust dynamics, and additional environmental drivers directly into the training dataset. By enriching the training data with simulations under diverse wind scenarios, the model can learn sensitivity to wind direction and intensity, thereby improving its applicability to more realistic operational settings.

To further clarify this positioning, we have added the following paragraph at the end of the Introduction:

It is important to emphasize that the present study is intended as a proof of concept demonstrating that a ConvLSTM-based surrogate model can successfully learn wildfire spread dynamics with dynamic firebreak deployment. The objective is methodological validation rather than the development of a fully operational forecasting system. While wind and landscape variability are critical drivers of real wildfire behaviour, incorporating fully dynamic, high-resolution environ-

mental forcing would substantially increase the dimensionality of the training space and require the generation of a significantly larger number of CA simulations. The proposed framework is flexible and can be extended in future work by enriching the training dataset with diverse wind directions, gust dynamics, and additional environmental drivers, thereby enabling the model to learn sensitivity to these factors and improving its applicability to more realistic operational settings.

We have also added the following clarifying sentence to Section 2.1 to explicitly justify the wind simplification:

This simplification was adopted to constrain the dimensionality of the training space in this proof-of-concept study.

In the Conclusion section, we have moderated the wording to avoid overstatement and to better align the interpretation with the observed results, as shown below:

~~*Despite these challenges, the model performs well, accurately capturing wildfire spread and firebreak behaviour across diverse scenarios. the model demonstrates consistent predictive capability within the tested scenarios and successfully captures key fire-firebreak interactions.*~~

Finally, we have expanded the Conclusion and Future Work sections to explicitly acknowledge the methodological scope and limitations of the present study and to outline directions for incorporating spatially variable wind forcing, detailed topographic information, and additional fuel characteristics in future work:

It is important to emphasize that the present work represents a methodological proof of concept rather than a fully operational wildfire forecasting system. The primary objective was to demonstrate that a ConvLSTM-based surrogate model can learn wildfire spread dynamics in the presence of dynamic firebreak deployment under controlled simulation settings. To maintain tractability of the training space, wind forcing was assumed to be spatially constant within each simulation, and landscape information was simplified. While these assumptions limit direct operational applicability, they allow for controlled evaluation of the surrogate modelling framework. Future work will focus on incorporating spatially and temporally variable wind fields, gust dynamics, detailed topographic information, and additional fuel characteristics into the training datasets. By enriching the diversity of environmental forcing scenarios, the model can learn greater sensitivity to key drivers of wildfire behaviour, thereby improving realism and extending its applicability to operational contexts.

Technical Comments

There is little to no discussion of the simulation results. All figures containing spatial information (maps) need improvement:

1. In Figure 1 the color selection makes it hard to interpret figures b, c, f and g. The natural color figures are also too small to see, “e” is of different size.

The colour maps in Figure 1 have been revised to improve interpretability and colour contrast, and the figure has been reformatted to ensure consistent panel sizing across all subfigures, as shown in figure 1.

2. In Figure 3 the simulation results are hard to see over a colorful background (especially when we are not sure what the colors denote, is it land cover?), either crop to the simulation zone, or utilize a blow-up window to show us the simulation results separately and in close up fashion. Alternatively, if you will not make any reference to the background, you can neutralize it with a filter, or other color selection so the BA and fire breaks are more visible, and preferably larger (same for Figure 8).

Following the reordering of figures, the original Figure 3 is now presented as Figure 4 in new manuscript. To improve visibility of the burned area and firebreaks, we have added a cropped and zoomed-in version with a neutralized background in Appendix B

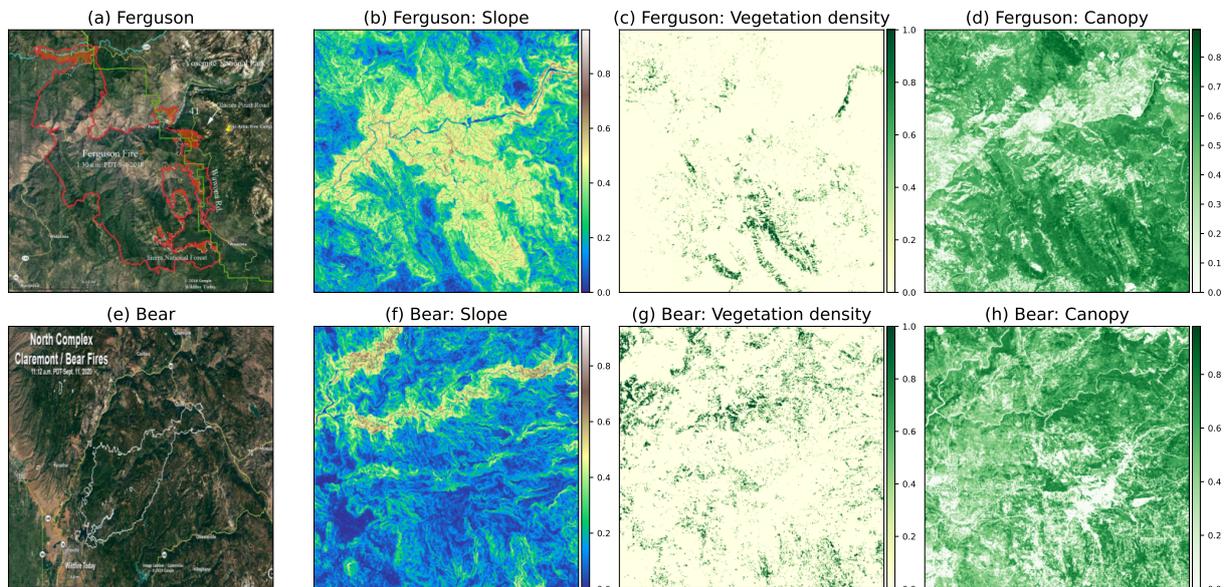


Figure 1: Updated figure 1.

(Figure B1). The caption of Figure 4 has been updated accordingly to reference this enhanced view.

A cropped and zoomed-in version is provided in Appendix B (Fig. B1), where the background is neutralized to emphasize the active fire scene and associated fire spread patterns.

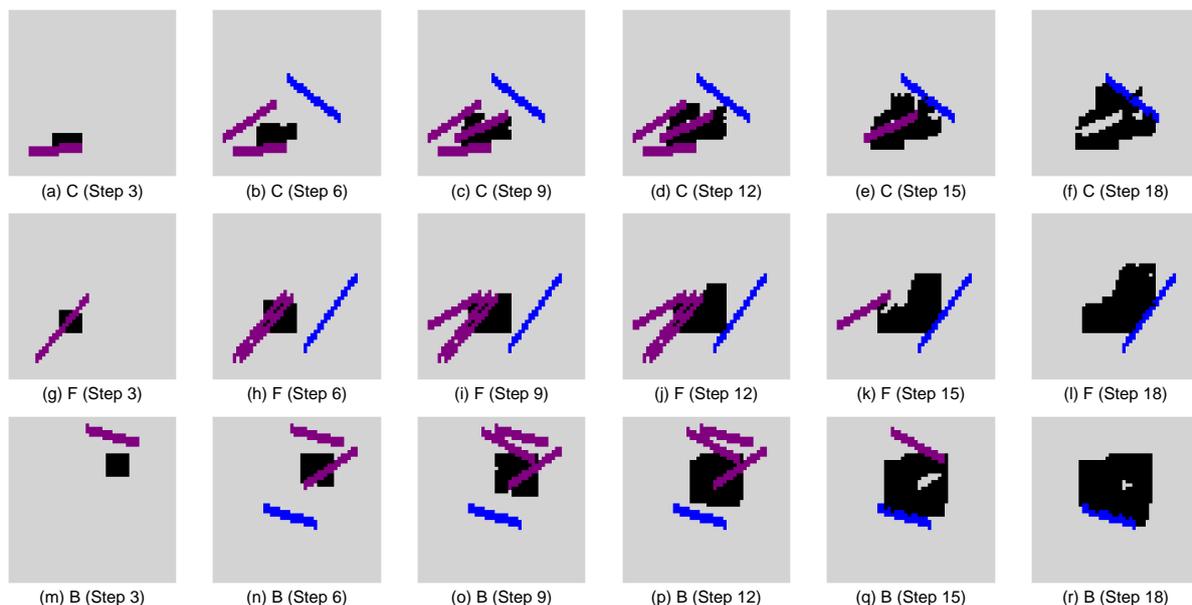


Figure 2: New figure B1. Cropped and zoomed-in view of Figure 4 in new manuscript, with the background neutralized to highlight the active fire scene and fire spread dynamics.

Similarly, for Figure 8 we have added a cropped and zoomed-in version with a neutralized background in Appendix B (Figure B2) to improve clarity of the fire spread patterns and error regions. The caption of Figure 8 has been revised to reference this additional visualization.

A cropped and zoomed-in version is provided in Appendix B (Fig B2), where the background is neutralized to emphasize the active fire scene and associated fire spread patterns.

The figure 3 shows the new figure B2.

For completeness and consistency, we have also added cropped and zoomed-in versions

of the additional test examples in Appendix B (Figures B4, B6, and B8), corresponding to Figures B3, B5, and B7, respectively. These enhanced views improve visibility of the active fire scene and firebreak structures across all test cases.

3. There is a Figure 4, but there is no reference to it in the text. I would have preferred it was introduced towards the beginning of the methodology when the reader is trying to visualize how the experiment runs. May be a few sentences before the detailed explanations of CA and ConvLSTM models introducing the workflow, and referencing the figure would also help understand the grid structure better, through actual visualization.

Thank you for this helpful suggestion. We have moved the workflow figure earlier in the manuscript and explicitly introduced it at the beginning of Section 2. The figure is now presented as Figure 2 in new manuscript and is accompanied by a new paragraph:

Figure 2 provides an overview of the data generation, model training, and prediction workflow used in this study. First, a CA simulator is employed to generate spatio-temporal wildfire propagation data under various environmental conditions and firebreak deployment scenarios. These CA-generated fire spread sequences, together with firebreak information, are then used to train a ConvLSTM-based surrogate model. During inference, the trained ConvLSTM model takes a sequence of previous fire states as input and predicts subsequent wildfire evolution with significantly reduced computational cost. This workflow illustrates the relationship between the physics-inspired CA simulator and the data-driven ConvLSTM model, and clarifies how the two components are integrated in the proposed framework.

4. In Section 3.2 the authors compare the speed of a CPU run simulation to a GPU run simulation, which will produce slower results. To be able to compare both simulations head on, they should be run on the same set up. CA model may run slower but from the rate quoted here it is unclear how much of it comes from the machine how much of it from the model’s own performance. Therefore a “250x” expression should be re-evaluated.

Thank you for this important comment. We agree that comparing a CPU-based simulation to a GPU-based simulation requires clarification.

Our intention was not to suggest that the ConvLSTM architecture is intrinsically $250\times$ faster than the CA model independent of hardware. Rather, the reported speedup reflects a practical deployment comparison between: a GPU-accelerated deep learning model (ConvLSTM), and a CPU-based Cellular Automata (CA) simulator.

To address the reviewer’s concern, we additionally benchmarked the ConvLSTM model on CPU under the same spatial resolutions. The results (shown in the figure 7 included in this response letter) indicate that when executed on CPU, the ConvLSTM model is substantially slower than the CA implementation, particularly at higher resolutions. This is expected because convolutional neural networks involve dense tensor operations that are highly optimized for GPU parallelization but are computationally expensive on CPU architectures [2, 3]. Deep learning frameworks such as PyTorch are explicitly designed to leverage GPU acceleration. As described in [1, 4], the framework is engineered for high-performance tensor computation on parallel hardware. Therefore, evaluating ConvLSTM on GPU reflects its intended operational setting.

We have revised Section 3.2 to clarify that:

The reported $250\times$ speed-up corresponds to a comparison between a GPU-accelerated ConvLSTM model and a CPU-based CA implementation, reflecting realistic deployment conditions. When executed on CPU, the ConvLSTM model is computationally more demanding than CA, as expected for convolutional neural networks without hardware acceleration.

This clarification ensures that the speed comparison is interpreted correctly and avoids attributing the performance gain solely to model architecture. We appreciate the reviewer’s comment, which has helped us present this aspect more transparently.

5. In the simulation results shown in Figure 8, there is a varying degree of false negative and positives among the three text examples. Ferguson fire being the smallest whereas Chimney Fire showing several (Bear Fire also). It is expected for a model that is trained on model data to exponentially over/under predict overtime, however the difference among the three test cases could have been better explained in the text. Also, I would expect to see a ratio timeseries (along with the map demarcations) so it is easier to interpret accurately. In an attempt to explain the false positive/negatives, a mention of landscape data limitation is mentioned here and wind speed, but the reader would appreciate a more in-depth explanation/discussion. Also “Despite these challenges, the model performs well, ..” is a bit of an overstatement given the results, toning that down may help meet expectations.

To address this concern, we have added the following paragraph at the end of Section 3.1 to provide a more detailed explanation of the differences in false positive and false negative patterns among the three wildfire cases:

The magnitude and spatial distribution of false positives and false negatives differ among the three wildfire cases, reflecting intrinsic differences in fire behaviour across landscapes. The Ferguson 2018 case exhibits relatively smaller error regions, whereas the Chimney 2016 and Bear 2020 cases show more pronounced deviations. This variation is likely associated with differences in vegetation density, fuel continuity, and wind-driven spread intensity across the study areas. Fires that propagate more rapidly or exhibit stronger directional bias due to wind forcing generate more complex and rapidly evolving fire fronts, which amplify autoregressive prediction errors over successive iterations. In contrast, slower or more spatially constrained fires produce more structured propagation patterns that are easier for the surrogate model to approximate.

In addition, because the ConvLSTM model does not explicitly incorporate detailed topographic features or spatially variable wind fields as dynamic inputs, the model must implicitly learn these effects from the CA-generated training data. This limitation may contribute to the larger discrepancies observed in landscapes where fire spread is more sensitive to environmental heterogeneity. Future work could integrate higher-resolution environmental forcing and real-time satellite observations through data assimilation strategies to dynamically correct prediction drift and improve long-term stability.

Regarding the suggestion of including a ratio time-series, the temporal evolution of prediction errors is already quantified through the Relative Prediction Error (RPE) and Structural Similarity Index Measure (SSIM), which are presented as time-series in Fig. 7. These metrics provide a quantitative assessment of the mismatch between predicted and simulated fire spread over time and complement the spatial error maps shown in Fig. 8.

And we have moderated the wording to avoid overstatement in the conclusion part, as shown below:

~~*Despite these challenges, the model performs well, accurately capturing wildfire spread and firebreak behaviour across diverse scenarios. the model demonstrates consistent predictive capability within the tested scenarios and successfully captures key fire–firebreak interactions.*~~

6. In sum, the authors undertake an important task, including fire breaks (and their efficiency) in fire propagation simulations. Among the pros the work’s easy reproducibility tops the list due to a clear methodological break down. However, these series of experiments are limited in capacity since they fall short of considering wind (speed and

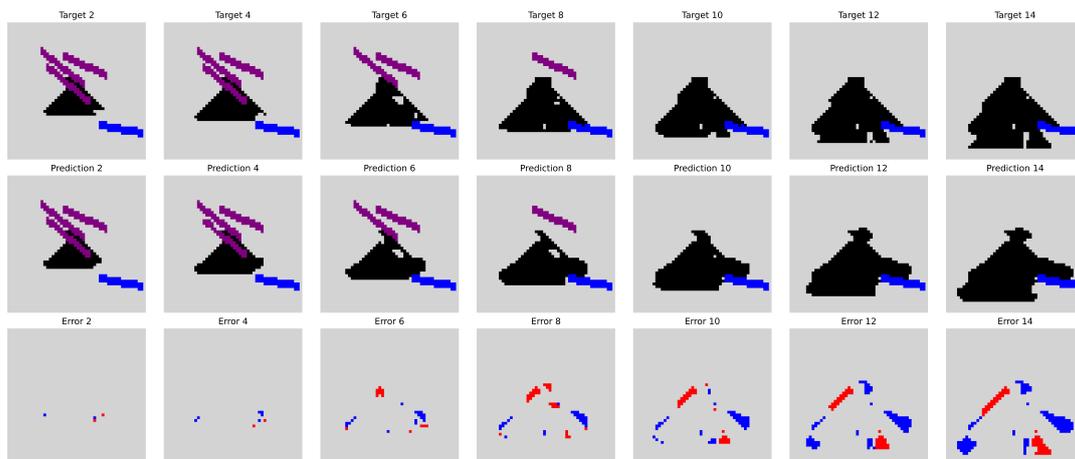
direction) as well as topography. The manuscript has room for improvement, especially through a solid discussion of results.

Thank you for this balanced and constructive summary. We agree that the current experiments adopt simplified wind forcing assumptions and do not yet include explicit topographic inputs in the ConvLSTM model, which constrain direct operational realism. To address this, we have added a dedicated discussion paragraph at the end of Section 3.1 clarifying these methodological choices and their implications. We have also strengthened the Conclusion section to emphasize that the present study represents a proof-of-concept framework and to outline future work incorporating spatially and temporally variable wind fields, gust dynamics, and explicit topographic inputs.

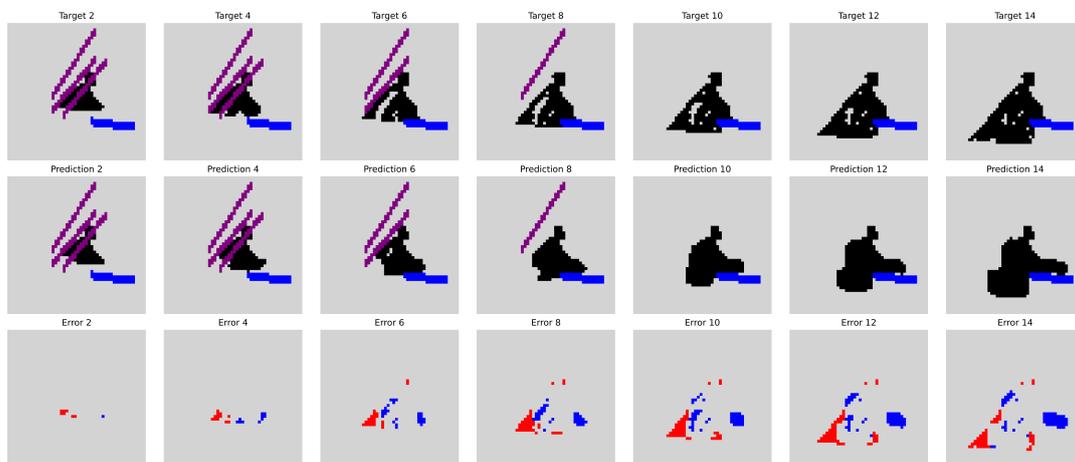
While the results demonstrate strong predictive capability within the tested configurations, it is important to recognize the methodological limitations of the current framework. Wind speed, wind direction, and slope effects are incorporated within the CA simulations to generate the training data. However, wind forcing was assumed to be spatially constant within each simulation, and detailed topographic information was not explicitly provided as dynamic input channels to the ConvLSTM model. In real wildfire events, spatially heterogeneous wind fields, gust dynamics, and terrain-induced flow effects can substantially alter fire spread behaviour. These simplifications reduce the dimensionality of the training space and allow controlled evaluation of the surrogate modelling approach, but they also limit direct operational realism. Future studies should incorporate higher-resolution meteorological inputs and explicit topographic features as model inputs to better represent complex fire–environment interactions.

References

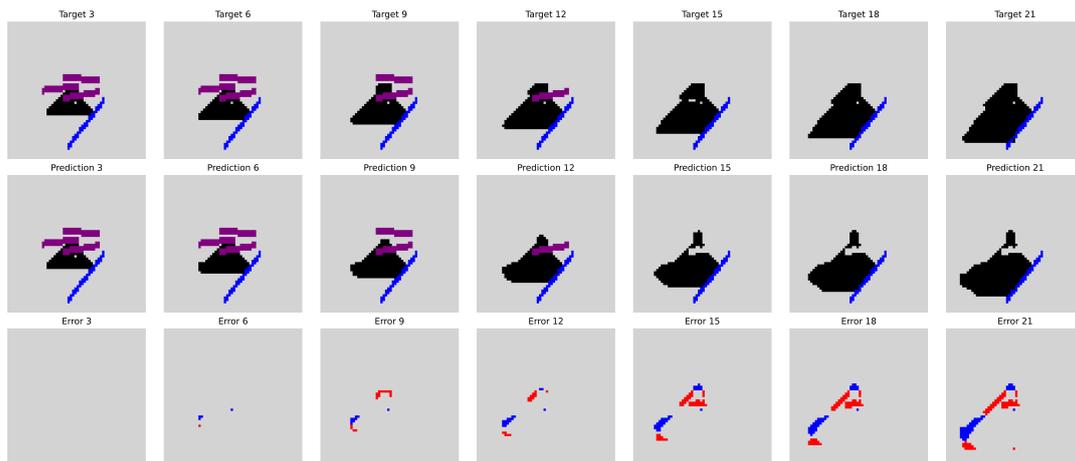
- [1] Torch-tensorrt. <https://docs.pytorch.org/TensorRT/>.
- [2] M. Boukili, N. Jebbor, K. Errajraji, and M. Elbathaoui. Cnn model acceleration and optimization: Cpu and gpu performance analysis and evaluation for multiclass classification. Available at SSRN 5325963.
- [3] D. Gyawali. Comparative analysis of cpu and gpu profiling for deep learning models. *arXiv preprint arXiv:2309.02521*, 2023.
- [4] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.



(a) Test example: Bear 2020

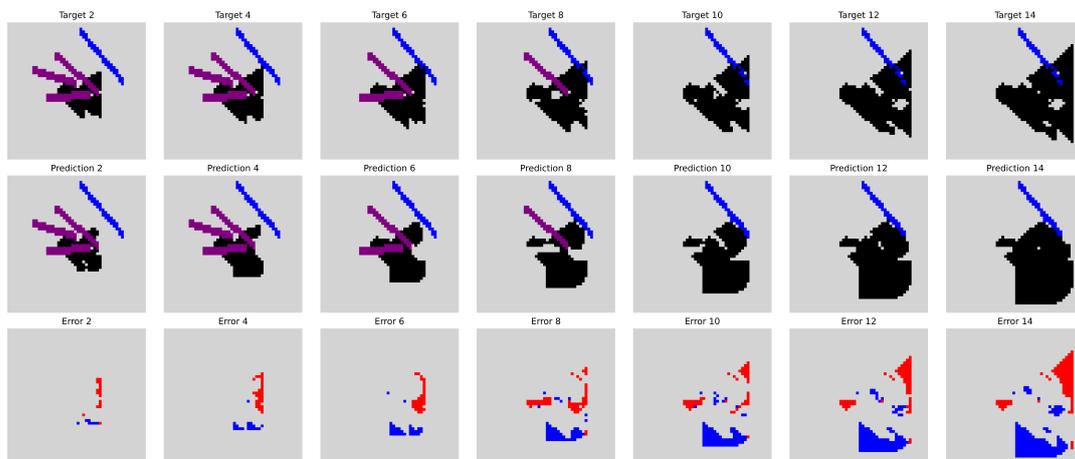


(b) Test example: Chimney 2016

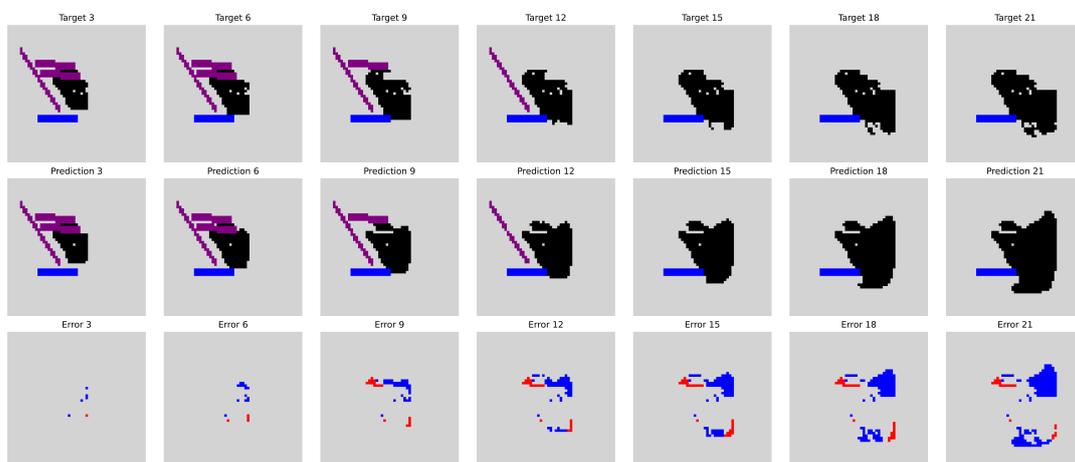


(c) Test example: Ferguson 2018

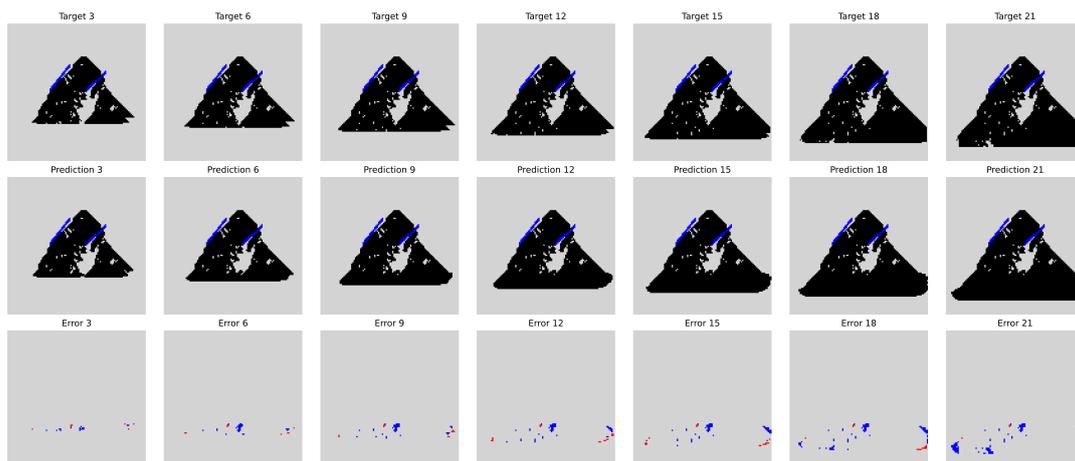
Figure 3: New figure B2. Cropped and zoomed-in view of Figure 8, with the background neutralized to highlight the active fire scene and fire spread dynamics.



(a)

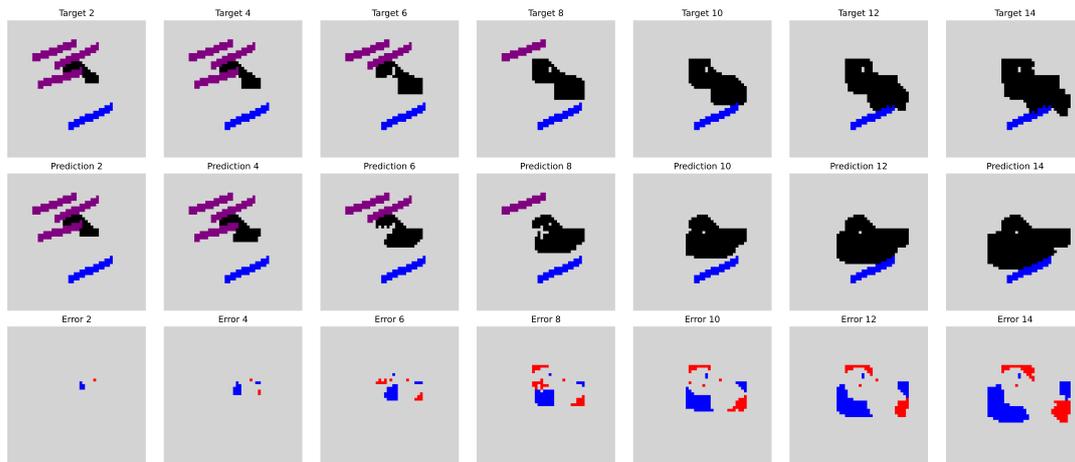


(b)

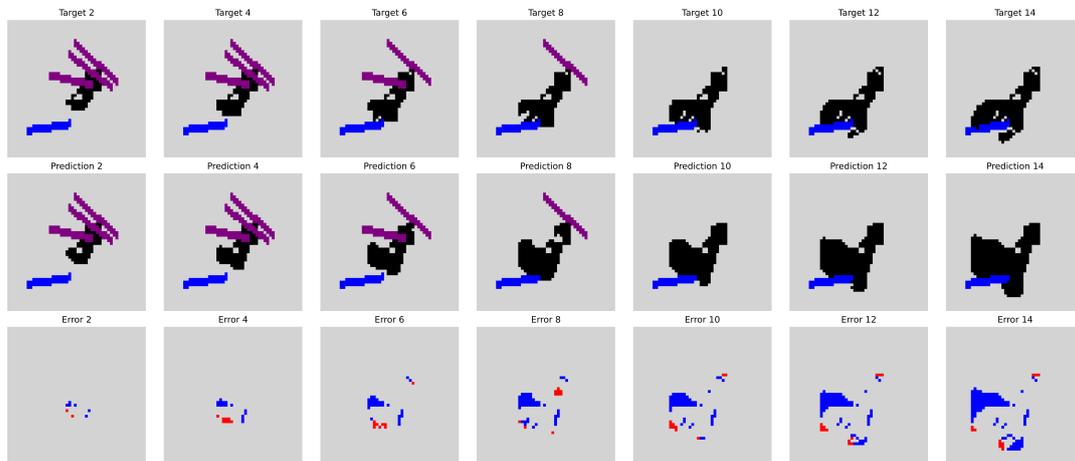


(c)

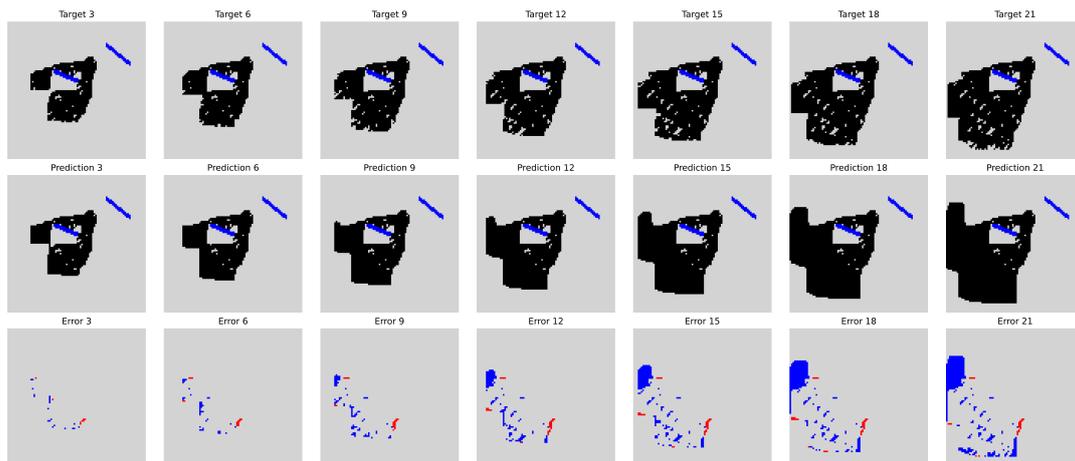
Figure 4: New figure B4. Cropped and zoomed-in view of Figure B3, with the background neutralized to highlight the active fire scene and fire spread dynamics.



(a)

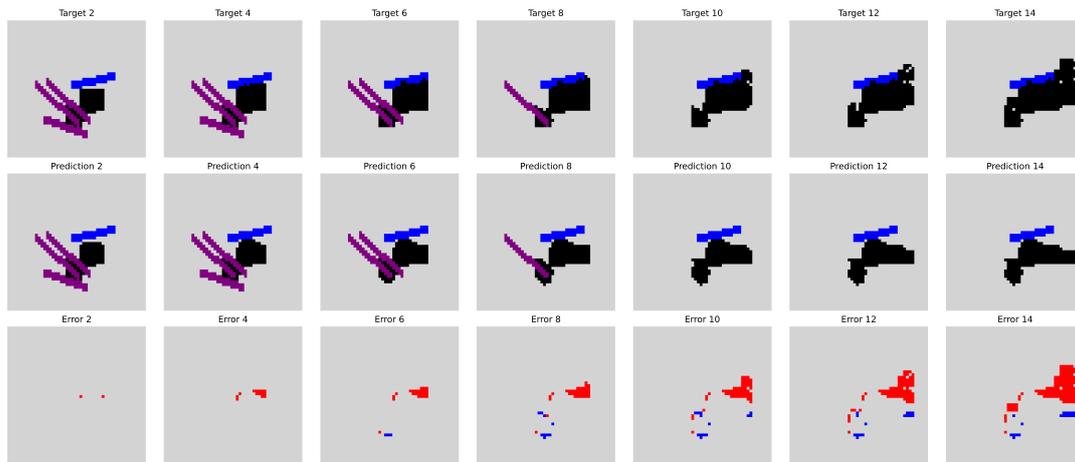


(b)

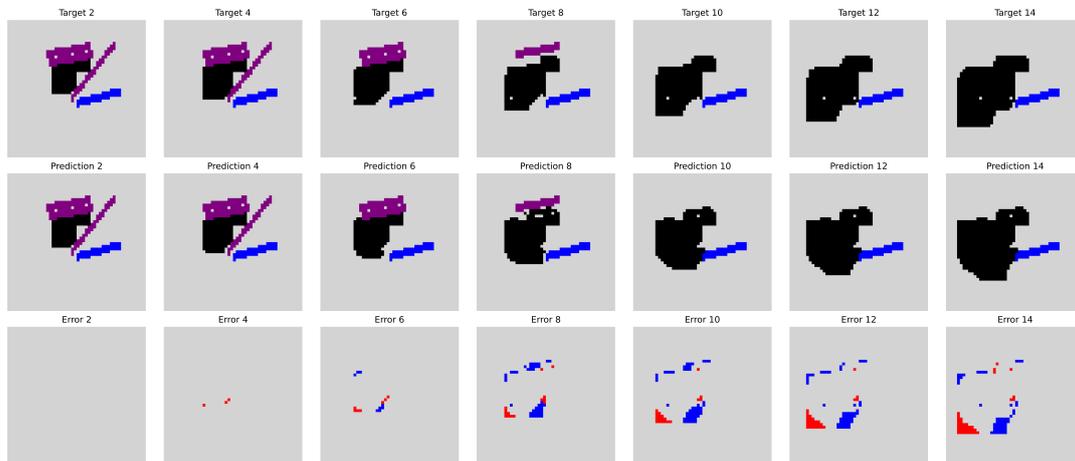


(c)

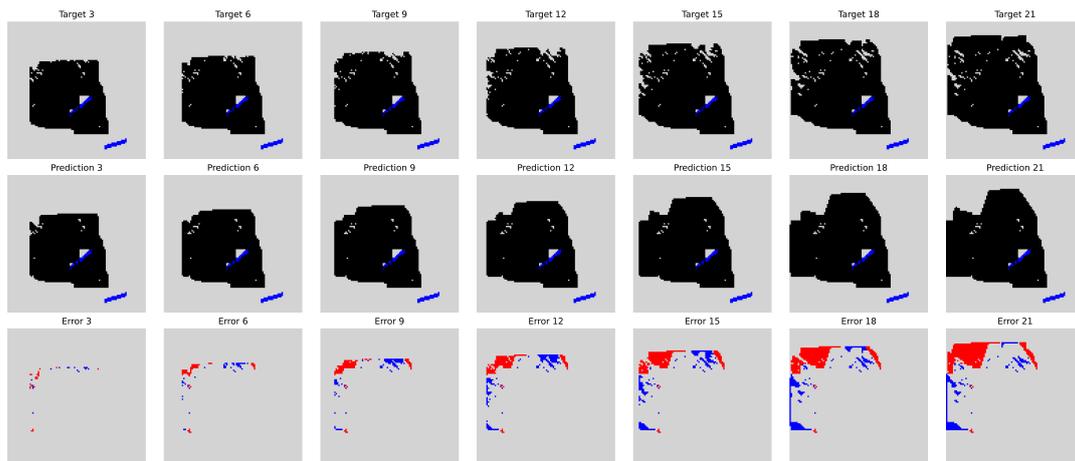
Figure 5: New figure B6. Cropped and zoomed-in view of Figure B5, with the background neutralized to highlight the active fire scene and fire spread dynamics.



(a)



(b)



(c)

Figure 6: New figure B8. Cropped and zoomed-in view of Figure B7, with the background neutralized to highlight the active fire scene and fire spread dynamics.

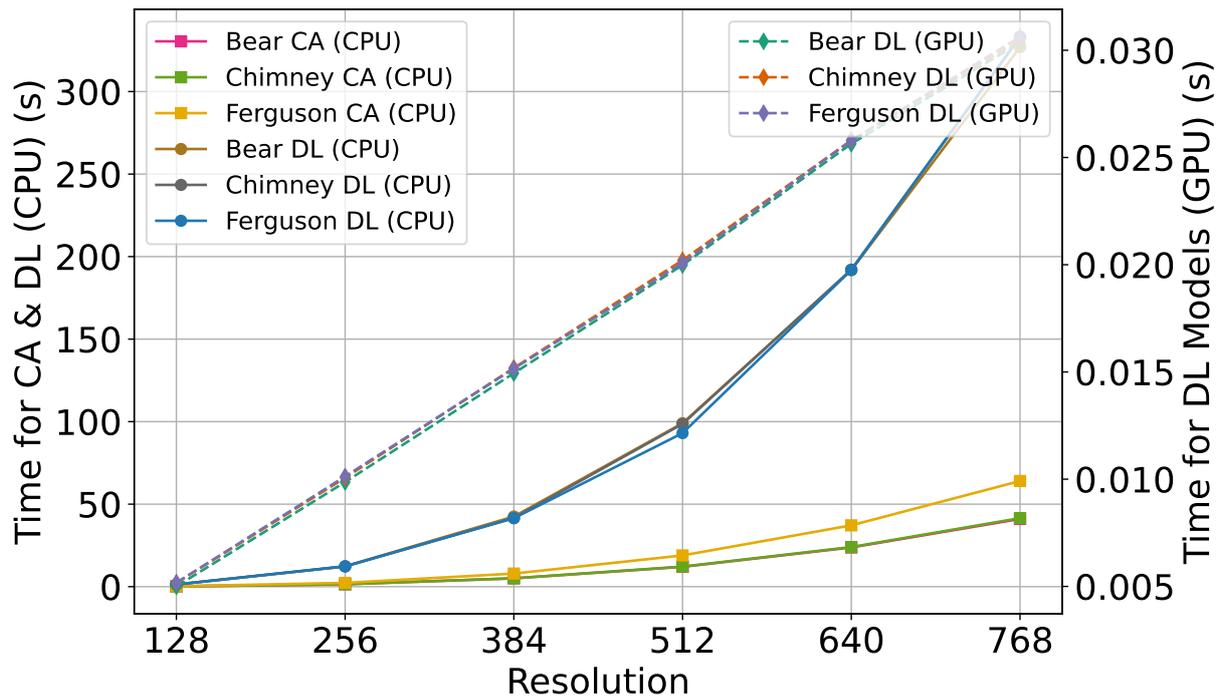


Figure 7: Runtime comparison of the CA model and the ConvLSTM (DL) model across increasing spatial resolutions (128×128 to 768×768). The left axis shows execution time (seconds) for CA and ConvLSTM when both are executed on CPU. The right axis shows execution time (seconds) for ConvLSTM executed on an NVIDIA A100 GPU. Results demonstrate that ConvLSTM inference is computationally expensive on CPU due to dense convolutional tensor operations, while GPU acceleration significantly reduces runtime through parallel processing. The previously reported “250 \times ” speedup refers specifically to the comparison between GPU-accelerated ConvLSTM and CPU-based CA under practical deployment settings.