Overall:

The paper presents a timely contribution to the rapid pluvial flood mapping using machine learning. The main novelty lies in the integration of larger "contextual terrain" information with high-resolution local "patches" (context-aware data-driven model), based on concepts from geospatial image segmentation. The methodology is solid and well-researched with multiple case studies (Zurich, Luzern, Singapore), different topographies, DEM resolutions, and rain events with varying return periods (2-100 years). "Transfer learning" from one terrain to another and a parsimonious retrain for new catchments indicates strong generalization potential. The results are well illustrated, using both visual and statistical criteria to evaluate the performance for unseen test data. However, some sections require clarification or additional details to strengthen the paper.

General/structural comments:

1. **L46-49**: You describe the need for rapid flood mapping mainly due to the long computational times of hydrodynamic models. To enhance this perspective, consider mentioning other issues that machine learning models can address, such as calibration flexibility and handling uncertainties in large or complex catchments. Also, acknowledge that while advances in computational power can reduce time for hydrodynamic models, they still don't match the speed and parsimony of machine learning for rapid flood mapping.
2. **L63**: Please spell out "CNN" as "Convolutional Neural Network" the first time it is mentioned for readers unfamiliar with the acronym.
3. **Section 2 and Fig. 2**: You reference Fig. 2 only once in Section 2. Some references to details in Fig. 2 in relevant parts of Section 2 would enhance their presentation. Ensure that the main figure and its caption in the text are as complete as those in the supplementary materials.
4. **L125**: Fix the formatting error for the subscript "norm". $S_i = \frac{P_{i,acc}}{P_n orm}$
5. **L203**: Clarify which model you are referring to in "the model to generate flood maps" (is it the proposed model or WCA2D?). Be consistent in using "generation" for target data preparation and "simulation" for the outputs of the data-driven model.
6. **L225**: Clarify that "18 1-hour uniform rainfall" refers to spatially uniform rainfall.
7. **L227**: Specify that "shapes" refers to hyetograph shapes.
8. **Table 1**: This table seems incomplete or needs redesign. Additionally, ensure it is referenced in the text (it currently isn't).

Figures:

1. **Fig. 5**: It's difficult to understand the transparency codes without reading the caption. Add a description, such as "(lighter color for shallower depth)" at the end of the text above the pie charts. Also, mention the case study in the caption, or in the title similar to Fig. 7.

2. **Fig. 7**: Apply the same suggestions as for Fig. 5 regarding the transparency. Additionally, use similar y-axis limits for Figs. 5 and 7a-b for easier comparison.
3. **Fig. 8 and Section 6**: Introduce the heatmap (Fig. 8) first and use it to strengthen your justification for choosing the P31-2 rain event for transfer learning.

Specific clarification:

1. **L248-252 (CSI Calculation)**: Clarify how "correctly identified cells" are defined, particularly whether this applies only for depths < or > 0.1 m. Given the high CSI values (0.98), it seems no threshold was applied for significant differences between target and simulated depths, but this should be briefly clarified.
2. **RMSE and CSI Values**: Summarize RMSE and CSI values across all case studies and events in a table for transparency and easier comparison.
3. **$RMSE_{0.1}$**: Define this index when first introduced (L333?).
4. **Transfer Learning (Section 6)**: The section title "terrain adaptation" is a bit misleading. Consider renaming it to "Improving generalizability via transfer learning and parsimonious retraining" to better reflect its focus.
5. Additionally, explain your choice of using only one rainfall event (P31-2) earlier in the section, building on Fig. 8 to justify this selection.

Discussion enhancements:

1. **Discuss the potential incomparability** across case studies since rainfall ranges differ by location, but seemingly you used same events from Zurich for Singapore. Singapore may receive more intense rainfall over longer period, resulting in higher water depths for larger areas (due to flatter terrain) and potentially higher modeling errors as you suggested for higher water depth ranges. This opens opportunities for further studies on transfer learning for various rainfall lengths and more diverse intensities. If my assumptions are wrong, you could possibly discuss based on right assumptions.
2. **Generalization Discussion**: Consider adding a reflection on combining hydrodynamic models and AI, or training models based on actual flood events and measured water depths. You could also mention the potential of integrating in-situ and remote sensing data to improve model performance when trained against actual observations.
3. **Hybrid AI and Hydrodynamic Models**: As your model uses multiple data-driven methods, a brief discussion on how future work could benefit from hybrid approaches—where machine learning augments traditional hydrodynamic models—would strengthen the paper’s relevance for broader applications. If there is a need for such approach, etc.