Review of “A Global Classification Dataset of Daytime and Nighttime Marine Low-cloud Mesoscale Morphology Based on Deep Learning Methods” by Wu et al. [MS number: essd-2024-536]

This study produces a global dataset of daytime and nighttime low-cloud mesoscale morphologies (categorized into six types) using a convolutional neural network through a combination of MODIS infrared radiance data and machine-learning-retrieved cloud optical thickness. Leveraging this novel dataset, the authors analyzed the day-night contrast in climatology, seasonal cycles, and cloud properties of cloud morphologies. One of notable findings is the significant diurnal variation in the occurrence frequency of closed MCC and suppressed Cu. The primary contribution of this work lies in the generation of nighttime low-cloud morphology data, which complements the well-established daytime morphology datasets from prior studies. This advancement would inspire and enable more downstream research like understanding the diurnal cycle of cloud morphology and cloud-longwave-radiation-climate feedback. The manuscript is overall well-written and well-organized, with nice presentation of figures. However, my major concern pertains to the limitations in the model’s training and validation processes, which could impact the dataset’s reliability. Addressing these issues would significantly strengthen the study’s contribution to the marine low-cloud research community. I’d like to recommend a major revision before this manuscript is considered for publication in ESSD.

Major comments:

1. One of my primary concerns is the validity of applying a regionally trained deep learning (DL) model to global predictions. In this study, the authors developed their model using data from the SEP region only and then applied it to generate a global dataset. While the model demonstrates relatively high prediction accuracy over SEP (Figure 3), it is unclear whether this performance extends to global applications. Regarding this issue, the authors should first clarify the rationale for selecting SEP as the training region rather than using a global or other regional dataset. Was this choice subject to the limited availability of the data, or is there a similarity in morphology climatology between SEP and the global scale? If SEP is your best choice at the moment, it would be essential to evaluate whether using a regionally trained model for global predictions is reasonable. One approach to examine this would be to generate a global map of prediction accuracy for each cloud morphology type to check the model’s global performance. Additionally, the authors could examine the differences in the PDFs of thermal radiance, COT, and cloud morphology between SEP and the global dataset. A smaller difference or larger overlaps would indicate less extrapolation by the model, enhancing the credibility of the global dataset.

Similarly, the authors would have to be careful when extending the daytime-trained model to nighttime predictions, as this may also introduce potential extrapolation issues. The authors provided only a single example to illustrate the model’s success at nighttime, which is insufficient to establish its statistical reliability. To address this concern, additional cases should be analyzed to validate the model’s nighttime performance. Alternatively, examining the differences in the PDFs of thermal radiance, COT, and cloud morphology between daytime and nighttime could help assess the extent of extrapolation and ensure the robustness of the predictions.

2. Regarding the model training, validation, and testing, the data-splitting strategy is unclear. For instance, was the dataset split randomly or manually into the 6:2:2 ratio? Furthermore, the validation method used to assess the model's predictions has not been described. The authors should clarify these aspects to improve the robustness of their results.

3. Given the critical role of cloud morphologies in Earth's radiation budget, the authors could consider including a climatological analysis of shortwave and longwave radiation at the TOA for the six cloud morphology types. Adding such an analysis would significantly enhance the insights and scientific value of this study.

Minor comments:

L30: longwave warming effects are more significant for high clouds, which might not be so for low clouds.

L67: What is the major difference between the six-type classification of this study and the four-type one here?

L81: Do you mean the decline in the *long-term* trend?

L83: "how much they contribute to … remain unclear" to "how nighttime cloud cover varies under different cloud morphology types remain unclear."

L90: Please clarify the temporal and spatial resolution.

L97: "created" to "driven"

L119: Please clarify the temporal resolution of the training dataset.

L121-122: Have you excluded middle clouds (i.e., those situated between 3 and 6 km)? These clouds are prevalent over midlatitude oceans, and they also contaminate low cloud observations.

L174: Please clarify the level of the divergence used.

L199: I'd suggest labeling the input variables (three channels and COT) and the output variables (six cloud morphology types) in Figure 2a to improve its clarity and readability.

L210: It looks like the improvement is limited. Have you examined the COT retrieval uncertainty? If it is greater than the improved accuracy, it would be unnecessary to include the COT into the predictors.

L210: Typo: "Yuan et al. (2020)_due to" to "Yuan et al. (2020) due to"

L212: Which is it relative to?

L219: "clustered Cu" to "clustered Cu or closed MCC"

L250: "n denotes" to "with n denoting"

L305: "its seasonal variation" to "the peak in summer"

L332: do you mean "decrease by 2 microns *on average*"?

L333: Please clarify whether the LWP mentioned here represents the in-cloud value or the grid-box mean value.

L349: Why is there a westward shift at night? Also, for stratocumulus clouds, LTS is usually higher at night. Why does it decline for closed MCC at night?

L351: It would be more interesting to discuss their physical reason.

L367: Why are the results shown here only for SEP, while Figure 10 presents global results?