We thank the reviewers for their in-depth comments and feedback as it provided great clarity for additional changes that need to be addressed in our work. Below are our responses to the comments and questions.

Reviewer 1:

Comment:

The manuscript is well written and provides application of machine learning to up-looking ground-based lidar.

The authors should also look at the very highly relevant publication: Fuller et al., (2025) "Using multitask machine learning to type clouds and aerosols from space-based photon-counting lidar measurements," Remote Sensing, doi: 10.3390/rs17162787. This recent publication is highly relevant and seemingly very similar to the submitted manuscript (i.e., using U-Net to perform cloud/aerosol sub-typing). Given the August 2025 publication date, the authors probably had not seen this paper before submitting theirs, but they should at least be aware of this publication as it does take ICESat-2 analysis beyond just the binary cloud-aerosol discrimination (a shortcoming specifically noted in lines 72-73).

Response:

We thank the reviewer for pointing out the relevant work by A. Fuller et al. (2025), Using Multitask Machine Learning to Type Clouds and Aerosols from Space-Based Photon-Counting Lidar Measurements. Indeed, we were not aware of the aforementioned paper during the initial submission of our work. We have added a citation and a discussion of this paper to our introduction/literature review.

While both studies utilize U-Net-based deep learning architectures to enhance Lidar retrievals, our approach differs fundamentally in its training objective and the physical capabilities of the resulting model. A. Fuller et al train their model using CATS L2O operational products, which are derived solely from CATS Lidar measurements. Consequently, their model, while improving resolution and detection of tenuous layers, remains constrained by the physical limitations of Lidar signal attenuation as it cannot be trained to classify atmospheric features where the Lidar signal is fully extinguished. In contrast, our model is trained on a single consistent labeling schema aligned with Cloudnet and PollyXT standards, which combine lidar and radar data. Because the radar penetrates optically thick layers that fully attenuate the Lidar signal, our ground truth contains "invisible" information (from the Lidar's perspective). This allows our model to learn contextual correlations and infer cloud and precipitation properties above the Lidar attenuation limit, effectively mimicking a Lidar-Radar synergy using only Lidar inputs. Thus, we have demonstrated that Cloudnet-like classification can be approximated using only a single lidar system. This represents a methodological and operational advancement beyond previous approaches. Revisions were added to the introduction and discussion sections.

Reviewer 2:

Major Comments

Major Comment 1:

Comment: The authors strongly emphasize that lidar-based models cannot observe beyond clouds due to complete attenuation, yet their own approach appears to do so (based on the emphasis in the introduction). However, by default, PollyXT cannot retrieve optical properties above clouds. How can this be predicted? If the lidar signal is fully attenuated at 5 km, how is the region between 5 and 22.5 km classified?

Response: We thank the reviewer for raising this important point regarding the observational limits of the PollyXT lidar and the precise objectives of our model. We fully agree that, by design, the lidar cannot provide direct observational information in regions where the signal is fully attenuated. Because of this physical limitation, a 1:1 reconstruction of the radar signal is not the actual goal of our approach, as this is likely impossible. Instead, the focus of our neural network is to extract structural and contextual hints to infer the presence of certain atmospheric objects, even when direct observability is lost. For example, while objects like dense precipitating cloud systems are not directly visible by lidar in their entirety, they leave distinctive traces in the lidar signal, such as falling precipitation profiles or specific cloud base geometries. The neural network exploits these traces, along with ancillary thermodynamic variables (temperature and pressure), to strongly constrain the likelihood of specific atmospheric states. These hints are highly valuable for cloud screening, cloud statistics, and general scene interpretation.

Consequently, classifications produced above the altitude of complete attenuation should not be interpreted as physical observations, but rather as model-based inferences. When the lidar signal is fully attenuated at, for example, 5 km, the region between this altitude and the upper model limit (22.5 km) is classified based on statistically learned and physically consistent patterns derived from the training dataset. This inference is primarily relevant for upper-tropospheric ice clouds and reflects climatological and thermodynamic coherence in vertically resolved atmospheric profiles. This behavior is conceptually similar to probabilistic classification approaches, such as those employed in Cloudnet-based frameworks, where classifications may persist beyond the limits of direct observability when sufficient contextual information is available.

To avoid ambiguity, we have revised the manuscript to explicitly state that classifications above complete lidar attenuation represent statistically inferred atmospheric states, guided by traces in the lower-level signal, rather than direct observations. Revisions were added to the Introduction and Section 4.2.3.

Major Comment 2:

Comment: Paragraph 43-51 may lead the reader to incorrect conclusions:

1. That the primary purpose of Cloudnet is to provide a detailed target classification.
2. That stand-alone lidar data may be sufficient to achieve this objective (potentially replacing Cloudnet).
   To my knowledge, Cloudnet serves multiple purposes beyond target classification. Moreover, even for this specific task, cloud radar offers greater capability to provide information within clouds (e.g., lower attenuation, information from Doppler spectra, etc.). It is somewhat paradoxical that the authors rely on the Cloudnet dataset to train their model while suggesting Cloudnet may be replaced for this task in the future. To what extent could algorithms be developed without the training database provided by Cloudnet?

Response: The reviewer makes an excellent point, and we apologize if our phrasing suggested an intent to render Cloudnet obsolete. This is absolutely not the case. Cloudnet serves a multitude of vital purposes beyond target classification, including the retrieval of continuous cloud microphysical properties (such as liquid/ice water content and effective radius), the continuous evaluation of numerical weather predictions and climate models, and the investigation of aerosol-cloud interactions. Furthermore, our AI model fundamentally relies on the high-quality, synergistic datasets generated by Cloudnet to serve as robust training targets. Our objective is not to replace Cloudnet, but rather to extract comparable cloud information from single-instrument observations due to the complexity and high cost of operating full Cloudnet setups. This AI application is intended to bridge the gap between the need for vertical cloud maps and the sparse availability of full Cloudnet stations, effectively extending Cloudnet-like target classification capabilities to sites equipped only with standard lidars. We have revised our introduction to better clarify this point.

Major Comment 3:

Comment: What is the impact of removing each of the input variables? Are all of them necessary? How do the performance metrics degrade when each input variable is removed? What about the variables derived from models (T and P)?

Response: We thank the reviewer for this important question. A comprehensive ablation study involving retraining the model while systematically removing each input variable was not conducted, primarily due to the computational cost associated with training the deep learning architecture and the large multi-year dataset used in this study. All input variables were selected based on established physical and observational considerations in lidar-based aerosol and cloud classification. In particular, the multi-wavelength attenuated backscatter and depolarization channels provide complementary information on particle size, shape, and phase, which are known to be critical for discriminating between aerosol types and cloud phases.

With respect to the ancillary thermodynamic variables, temperature (T) and pressure (P) primarily act as contextual information, constraining physically plausible cloud phase and altitude relationships rather than providing direct classification signals. As such, we expect that removing T and P would lead to a moderate degradation in cloud‑phase discrimination, particularly for mixed‑phase and ice clouds, while having a smaller impact on aerosol classification, which is dominated by optical properties.

Importantly, the strong aerosol classification performance is achieved primarily from the lidar optical signals whereas cloud classification benefits from the additional thermodynamic context, and is consistent with prior Cloudnet‑based methodologies.

We have clarified this point in the revised manuscript by explicitly discussing the expected role and necessity of each input variable, while noting that a full quantitative ablation analysis is a valuable direction for future work.

Major Comment 4:

Comment: It is stated that the lidar temporal resolution is 90 seconds up to 24 hours, and the vertical resolution is 37 meters up to 22.5 km. However, Cloudnet variables are not always at that resolution and are often only available up to 11-12 km. At which resolution is lidar-Cloudnet data compared? Which variables are used? I strongly recommend that the training dataset is shared in an open repository for the sake of reproducibility.

Response: We thank the reviewer for pointing out this necessary clarification. Code and datasets are already shared and available in an open repository, as described in the 'Code and data availability' section of the paper. Regarding the resolution mismatch between the lidar and Cloudnet datasets, the comparison and training are conducted at the finer lidar resolution (90 seconds, 37 meters). To achieve this without altering the categorical nature of the target variables, the Cloudnet data was interpolated onto the PollyXT grid using nearest-neighbor interpolation. This approach simply replicates the Cloudnet data onto the finer lidar grid, leaving the categorical values entirely untouched and avoiding any unphysical blending or averaging of discrete masks. We have updated Section 2 of the manuscript to explicitly describe this interpolation strategy.

Major Comment 5:

Comment: The statement regarding the synergistic combination of PollyXT and Cloudnet outputs is not clear enough to reproduce. This combination must be explained in detail: Which method is used for the target classification of the PollyXT? How are inconsistencies between the two algorithms solved? Are there any assumptions made?

Response: We thank the reviewer for the opportunity to clarify this methodology. The merging process is straightforward and relies on leveraging the distinct observational strengths of both instruments.

Specifically, the baseline target classification is derived from the PollyXT classification scheme described by Baars et al. (2017). We then apply a simple overriding rule: wherever the radar instrument detects cloud particles, the base mask is overwritten by the Cloudnet target classification mask (Illingworth et al., 2007).

The fundamental assumption guiding this approach is that radar provides superior detection of ice crystals and precipitation compared to lidar. Consequently, the lidar classification serves as a fallback, utilized only when the radar yields no signal. This methodology offers three distinct advantages:

1. Preserves aerosol data: The underlying lidar aerosol classification remains entirely untouched.
2. Enhances precipitation data: The precipitation signals are substantially improved by prioritizing the radar data.
3. Retains high-altitude ice clouds: High ice clouds, whose signals are often too weak for radar detection, are still successfully captured by the lidar fallback.

We have revised the manuscript to describe this merging process more explicitly and have ensured both Baars et al. (2017) and Illingworth et al. (2007) are properly cited.

Major Comment 6:

Comment: The replacement of missing values using a global mean (gap filling) may introduce bias. Signal attenuation in thick clouds or low signal-to-noise in pristine regions is physically meaningful, not just an "absence of measurement." This methodology must be justified and its consistency demonstrated.

Response: We appreciate the reviewer's concern regarding potential bias from global mean imputation. Because neural networks cannot mathematically process NaN values, numerical imputation was a structural necessity. To address the exact bias the reviewer mentions, and to prevent the model from confusing a physical measurement with an imputed "absence of measurement," we explicitly engineered a binary indicator feature. For each input feature, a corresponding binary mask was created that took a value of 1 if the original data was a NaN (and subsequently imputed), and 0 otherwise. This step provided the model with explicit information about the original data quality at each point, allowing it to learn the distinction between valid signals and attenuated regions. We emphasize this binary indicator step more prominently in Section 3.1.

Major Comment 7:

Comment: The use of a log(1+x) transformation for lidar-derived variables is noted, but the justification is insufficient. The claim that it leads to a Gaussian distribution is not demonstrated. The choice of log(1+x) over alternatives is not justified. It is unclear if this is applied to all features (e.g., Ångström exponent). The authors should provide statistical or distributional analyses over the full dataset rather than a single example.

Response: We thank the reviewer for pointing out that we did not demonstrate effectively the Gaussian-like effect a log transform has on the data, and we acknowledge that the justification for the transformation was brief. This transformation was strictly necessary because lidar signals, particularly backscatter coefficients, span several orders of magnitude. Without compressing this dynamic range, the disproportionate influence of high-intensity values would dominate the loss function, preventing the neural network from learning morphological features in lower-intensity regions. While standard z-score normalization was applied afterward, doing so without a prior log transformation on heavily skewed data leaves the network highly sensitive to extreme outliers. Unfortunately, due to the size of the dataset it is not possible to show the entire distribution of one feature. However, we will demonstrate the effect of the transformation on a larger sample of the dataset and provide a clear distinction of the features that underwent this transformation and those that did not.

Major Comment 8:

Comment: From the analysis (lines 226-228) and conclusions (lines 323-326), cloud classification performance is substantially lower than aerosol classification. Given this and lidar signal attenuation, it is difficult to justify the title: "Cloud Fields and Aerosol Classification."

Response: We thank the reviewer for their important comment and for carefully examining the reported performance metrics. We fully acknowledge that aerosol classification yields the highest overall skill, with recall values of approximately 95–97% across aerosol subclasses. However, the manuscript also demonstrates that the proposed model achieves physically meaningful performance for cloud‑field classification, particularly for ice‑phase clouds.

Specifically, upper‑level ice clouds are correctly identified with a recall of approximately 79%, as reported in the confusion matrix and further illustrated in the dedicated case studies. As shown, the model successfully infers the presence and vertical structure of ice clouds even at, and beyond, the effective lidar attenuation limit, addressing a well‑known limitation of elastic lidar observations.

The comparatively lower performance for liquid cloud classes is primarily driven by confusions between adjacent and physically similar categories (e.g., “water droplets” versus “likely water droplets”), rather than by systematic aerosol–cloud misclassification. This reflects the continuous nature of cloud microphysical processes and the difficulty of assigning sharp class boundaries, rather than a failure to detect or represent cloud structures.

Importantly, the manuscript does not claim equivalent classification difficulty or accuracy for aerosols and clouds but instead presents a unified framework capable of addressing both, while explicitly quantifying and discussing their differing levels of uncertainty. For these reasons, we believe the title accurately reflects the dual, but not symmetrical, capabilities of the proposed approach. We will make adjustments to the abstract to better make this point.

Minor Comments

Minor Comment 1:

Comment: When a single pixel is classified by both PollyXT and Cloudnet (e.g., intersecting aerosol layers/low clouds), which class takes priority?

Response: See response to Major Comment 5.

Minor Comment 2:

Comment: How are the mean and variance retrieved (by profile, image, or whole dataset)?

Response: The mean and variance are calculated globally, per feature, strictly on the training dataset alone. These calculated parameters are then frozen and applied to scale the validation and test datasets. We chose this global approach to prevent data leakage from the evaluation sets into the training process, and to ensure we preserve the relative, absolute magnitude differences between different atmospheric events (which would be lost if standardized per-image or per-profile). This is briefly stated in Section 3.1 of the manuscript.

Minor Comment 3:

Comment: Fig. 1: Report height in meters rather than pixels/bins.

Response: The figure was changed to reflect the height in meters.

Minor Comment 4:

Comment: Line 142: Figure 4 is mentioned before Figure 3.

Response: Figures were correctly numbered to represent their actual location in the paper.

Minor Comment 5:

Comment: I suggest showing the attenuated backscatter at 1064 nm alongside ground truth and model predictions (Figures 7, 8, and 9).

Response: Attenuated backscatter at 1064 nm exists for each of the case studies in the Appendix, however for easier visual assessment we will add it alongside the ground truth and model predictions.

Minor Comment 6:

Comment: What is the LWP (liquid and ice) for case studies 2 and 3?

Response: We have added the LWP chart for each of the case studies in the appendix.