This study presents a full-column aerosol and cloud classification using only standard lidar inputs (from PollyXT) by means of an AI approach trained with CLOUDNET database. From my point of view, this study is very interesting for the journal and the scientific community, but I have concerns related to the degree of explanation of the promising methodology. Since it is an innovative approach and the AI approach is relatively new, I would recommend a major revision before accepting its publication. I would suggest authors to consider the following comments and correct the found typos.

Major 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. Therefore, how can this be predicted? If the lidar signal is fully attenuated at 5 km, how is it classified the region between the 5 and 22,5 km?
- Paragraph 43–51 may lead the reader to incorrect conclusions: (1) that the primary purpose of Cloudnet is to provide a detailed target classification, and (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.). Indeed, it is somewhat paradoxical that the authors rely on the Cloudnet dataset to train their model, while at the same time suggesting Cloudnet may be replaced for this task in the future. This raises an important question: to what extent could algorithms be developed without the training database provided by Cloudnet?
- 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? And what about the variables derived from models (T and P)?
- Line 96-102: In line 96 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, the variables of the Cloudnet are not always at that resolution and is they are only available for up to 11-12 km. At which resolution is lidar-Cloudnet data compared? Which variables are used? This training dataset seems to be key in the process, so I would strongly recommend that the training dataset is shared (an open repository?) for the sake of reproducibility.
- Line 104-106: Linked to the previous question, I think the statement of “It was constructed synergistically combining outputs from PollyXT aerosol target classification and Cloudnet cloud and precipitation target classification” is not clear enough to reproduce the approach. As far as I know, training datasets are crucial in the framework of AI applications. Therefore, this combination shall be explained in detail because it is not an easy process. Many questions (such as which method is used for the target classification of the PollyXT?, How are inconsistencies between the two target classification algorithms solved?, Are there any assumptions made?) arise reading this paragraph. These questions are

important and need to be clarified as part of the methodology. (see also minor comment Line 103).

- Section 3.1: The replacement of missing values in the script uses a global mean. Despite I am not an expert on AI, this 'gap filling' may have an impact since it is not just 'absence of measurement' but signal attenuation in optically thick clouds or low signal-to-noise regions in pristine regions, both examples with a physically meaningful a clear physical meaningful. May this introduce a bias in the model? A detailed explanation shall be included in the methodology. Summarizing, the 'gap filling' methodology shall be specified and justified, and their consistency shall be demonstrated.
- Figure 2 and lines 119-124: The use of a log(1+x) transformation for lidar-derived variables is reasonable given their large dynamic range. However, the justification provided is not enough from my point of view. For example, the claim that it leads to an approximately Gaussian distribution is not demonstrated and may be overstated, as log-transformations do not guarantee normality in such data. The choice of log(1+x) over other alternatives is not justified, and it is unclear whether the transformation is applied to all features, including those that may not require it (e.g., Ångström exponent). Moreover, no empirical evidence is provided to show that this preprocessing step improves model performance. The reference to Box and Cox (1964) is also questionable without formal optimization. Finally, the example shown in Figure 2 is unclear whether it is representative of the overall dataset. The authors are encouraged to provide statistics or distributional analyses over the full dataset (rather than a single example) to support the general validity of the transformation.
- From the analysis (lines 226–228) and the conclusions (lines 323–326), it can be inferred that the performance of cloud classification is substantially lower than that of aerosol classification. This, together with the lidar signal attenuation within clouds, makes it difficult to justify the title of the paper, "Cloud Fields and Aerosol Classification".

Overall, a more rigorous and quantitatively supported justification of the methodological choices is needed. Providing a clear and well-founded rationale for these decisions would greatly benefit the scientific community, rather than presenting a sequence of steps that may appear arbitrary.

Minor comments:

- Line 103: In cases where a single pixel is classified by both PollyXT and Cloudnet (for example, aerosol layer / low clouds that intersect), which class is selected? Is any priority (PollyXT > Cloudnet or Cloudnet > PollyXT)? I think this is not specified.
- Line 125: How the mean and variance is retrieved? By profile? By image? For the whole dataset?
- Fig. 1: It would be advisable to report height in meters, rather than in pixels/bins.
- Line 142: Figure 4 is mentioned before Figure 3.
- Although the raw signals (pixel values) are shown in the Appendix, I would strongly suggest showing the attenuated backscatter at 1064 joint to ground truth and model prediction (Figures 7, 8, and 9), helping to understand the input-

output relationship. Which is the LWP (liquid and ice) of the case studies 2 and 3?

- Line 5: gap in cost of data acquisition -> gap in the? cost of data acquisition
- Line 9: high precision, recall and F1-scores -> high precision, recall, and F1-scores
- Line 18: gasses -> gases
- Line 21: Aerosol-cloud interactions (ACI), represent -> Aerosol-cloud interactions (ACI) represent
- Line 31-32: its different vertical distribution -> their? different vertical distribution
- Line 62: it infer the -> it infers
- Line 100: 532 nm , -> 532 nm,
- Line 116: occur -> occur