Reviewer 1

The authors have submitted an interesting manuscript that presents the principle and building of a artifical neural network (ANN) based algorithm and its results. The objective is the identification of the multilayer character of ice-over-water multilayer clouds from MODIS/AQUA measurements. The ANN is trained with information from the active sensors of the A Train. While several approachs and attempts to retrieve such information about multilayer clouds from passive sensors exist, and they are listed in the manuscript and used for comparison, multilayer qualification is still a quite innovative retrieval target.

Authors give arguments about the importance to get this single layer (SL) and multilayer (ML) qualifications in order to improve the estimate of radiative budget, and to assimilate with improved confidence informations about clouds in different applications.

Data and methodology are clearly presented as well as the results. The analysis of the results are interesting : recall of ML cloud detected as a function of total and upper layer COD, probability distributions of true and false SL and ML clouds as a function of upper layer COD, attempts to correct and validate the use of the neural network as a function of the viewing angle in order to exploit it over the entire MODIS swath, presentation of case studies, demonstration with inferred climatologies of ML cloud occurrence of the performance and value of the approach, operationnal considerations.

Overall, the performance of the algorithm appears robust, consistent and, even if the comparison is not so straightforward. the performance's scores seem higher that those of other algorithms. The statistical bias of the algorithm appears to lie between -5 and -10 % (ML occurrence), bias that disappears when studying monthly anomalies following the fact that the bias is latitude independent and certainly, not shown, timely independent. Two very valuable aspects of this algorithm are that is performs during both day and night and also above snow-covered surfaces.

This work and its results fit well in the scope of EGUsphere, topic Atmospheric sciences and AMT.

As a reviewer, I will call here for minor revisions of this manuscript.

Thanks for reviewing this paper and providing constructive comments for improvement. We hope the concerns have been addressed satisfactorily.

Some omissions bring some lack of clarity.

First, the neural network algorithm appears here a little bit as a black box. It is understandable, as it is, in a way, a black box, but some descriptions could have been turned differently.

The usage, two times, of ‘artificial intelligence’, to describe its operation, participates in this shortcut. The ANN works here thanks to adequate choice of inputs and outputs and trainings. So it is not by itself intelligent.

It is true that the ANN is not by itself intelligent. However, the artificial neural network is recognized as a form of artificial intelligence, albeit a somewhat primitive one having first appeared in 1954. However, to alleviate worries that the MLANN might be intelligent unto itself, we have eliminated the terms from this paper.

Authors don’t give two much technical details about how they design for their usage the NN, letting the reader find it in previous studies as Minnis et al (2016, 2019) and Sun-Mack et al (2017). And this technique seems to be in the continuity of the works of Sun-Mack et al (2017) and Minnis et al (2019). The current manuscript makes numerous times reference to these two previous communications. The first one is a conference paper, not necessarily a problem, but the communication is not very long and detailed. More problematic, the second one, Minnis et al (2019) is not given in the references ’page of the paper. So one should accept all what is said about this reference (and there are some lines as on lines 93 to 98, etc.) without having the possibility to read it … The reference should be given.

We have rewritten the Data and Methodology sections. Hopefully, this will clear up some of the problems. References have also been included.

Finally, the (too?) short description of the physical basis behind the usage of the ANN leads to some lack of clarity :

- concerning the contraint used for the training (top of page 8), the rationale is not given : why no temperature inversion between 273 and 253 K ? the rationale should certainly be added in order to better understand the applied filtering and the definiton of what are here selected as multilayer cloud situations, that would help in the appreciation of the difference between the presented results and the ones of the others references (MYD06 C6, POLDER, …)

The second condition has been rewritten as follows. We hope that clarifies the reasoning.

“(2) at least, one layer with extinction occurs at a height above the altitude corresponding to 253 K and no temperature inversion exists in the atmospheric layer between the altitudes corresponding to 273 K and 253 K. This constraint is used to eliminate the possibility of warm clouds occurring above the assumed ice threshold of 253 K.”

- It is said on line 325 that ‘The results represent a significant improvement over the previous formulation. Much of the increased accuracy is due to …’ but the rationale of it could be clearer.

We have rewritten that paragraph as follows. We hope it is more understandable now.

"These results represent a significant improvement over the previous MLANN formulation (Minnis et al. 2019), which only attained accuracies of 80.4% and 77.1% during the day and night, respectively over SF surfaces. Much of the increased accuracy is due to use of shorter CALIPSO horizontal averaging distances here. By employing CALIPSO averages over distances up to 80 km, Minnis et al. (2019) attempted to detect ML cloud systems that included many cirrus clouds having optical depths smaller than 0.2. Such clouds are difficult to detect with passive remote sensing even when they are single-layered. According to Yost et al. (2021), systems having $\tau_{CC} < 0.2$ account for ~42% of all ML clouds for CALIPSO data using HA $\leq$ 80 km compared to only 18% for HA $\leq$ 5 km. A majority of those low-optical-depth ML clouds were not detected in Minnis et al. (2019), resulting in lower accuracies. Typically, cloud identification or multilayered cloud detection methods that use CALIPSO for validation or training employ data with HA $\leq$ 1 or 5 km (e.g., Desmons et al., 2017; Marchant et al., 2020; Tan et al., 2022; White et al., 2021). By using that smaller averaging distance in this study, the fraction of CC ML clouds is ~25% less than that used by Minnis et al. (2019), but a larger portion of them is detected. Other sources for the improvement arise from utilizing additional input parameters, including those based on the 13.3-µm channel and $\rho_{1.38}$, and $\rho_{1.61}$ - $\rho_{2.13}$. Additionally, the assumption that all pixels having $\tau_{CM} < 0.5$ are automatically SL, regardless of the CALIPSO classification, probably removed some difficult but less important cases."

- On line 332 : 'improvements arise from using additional input parameters, including [IR and NIR] channels'. Also on line 333 : 'vertical profile of relative humidity'. Wasn't there a way to argue or illustrate more clearly, if not demonstrate, these added values ? (why and how much the vertical profile of relative humidity adds useful information ?).

Please see the 3rd paragraph in section 3.0 where we specify the contribution of the RH profiles.

Minor Questions :

- a basic question : why the choice of 50-70 neurons for the hidden layer is a good choice ?

See the addition to the first paragraph in the Methodology section.

"Only one hidden layer is used for this shallow neural network. It was found that a second layer yielded no significant increase in accuracy, but greatly increased processing time. The number of neurons in the hidden layer vary from 50 to 70 depending on the data category (e.g., snow-free daytime ice clouds). The exact number was determined by adding neurons until gains in accuracy ended."

- Concerning the choice of defining a cloudy pixel as ML when the MLANN output probability is higher than 0.5 :

What is the consequence of the threshold's choice ?

The accuracy decreases for values < or > 0.5 for ice < or > 0.55 for water. The difference in accuracy between using 0.5 or 0.55 as the threshold for water is 0.1% relative to the value at 0.50. Thus, we use 0.5 as the threshold for both. The paragraph in the new methodology section summarizes our reasoning (see next answer).

Have you thought about defining instead as output a probability between 0 and 1, as in Desmons et al (2017), with a definition of a binary threshold that would maximize the algorithm accuracy ?

We changed the statement to ,

"Output from the trained MLANN is a probability between 0 and 1 for each pixel. The latter value denotes certainty that the pixel includes ML clouds as defined here. For practical purposes it is necessary to select a threshold probability above which a pixel is designated as multilayered. A threshold value of 0.5 was chosen based on analysis of the accuracy of the results for probabilities between 0.3 and 0.60. The accuracies (risks) were found to be greatest (least) for thresholds between 0.50 and 0.55."

- a distinction is made between snow-free and snow-covered surfaces, which makes sense. Wouldn't it have been interesting to present performances with a distinction ocean/land ? (that would have shown a better performance during the night over land?

Training for land and ocean separately may gain additional accuracy, especially over land. However, one of the input values is surface type, land or ocean, which should lead to optimal performance over each surface. We can examine that in future studies.

- on line 270 : if CoS is called Single-layer Confidence, why not defining for consistency on line 268 PR as Multilayer confidence (CoM) ?

This is one of the terminology issues that has driven us and apparently Reviewer 2 to distraction. Desmons et al. (2019) used CoM, while Tan et al. (2022) used precision. We opted to use the Tan et al. (2022) nomenclature except for parameters they did not include.

- on line 270 : isn't the definition or equation wrong ? Isn't it instead False SL rate ? Or SS/(SS+SM). Values given in Tables seem correct ; but the definition is certainly wrong here.

Yes, the formula was not written correctly and has been revised.

- on line 511 : why is it 'not shown' ?

We were trying to keep the plots simple. We now show the full -70 to 70°. It is true they were not cancelling each other. The ice and water curves were mistakenly used to make the original statement. Thanks. The paragraph has been changed accordingly.

- From the results of Figure 5 : would there be an interest to plot the difference (night minus day) and show the capacity of MLANN to get right this difference ?

We have plotted it and found $R^2 = 0.88$ and included it as the new Fig. 5.

Minor comments or typo :

- line 88 : does the reference to Minnis et al (2023) exists ? It should be given.

Done

- line 108 : Venetsanopoulos

Not sure what is needed here. The name is spelled the same as in the paper.

- line 120 : equation with equal sign not to be cutted in two

Done

- line 315 : Table 3 instead of Table 2

Done

- sentence on line 368 and 369 : is it really two conditions on \tau_{CM} ?

Good catch. We corrected the second condition to reflect tauC, instead of tauCM.

- line 444 : Desmons instead of Desmond

Done

- line 446 : Marchant et al (2020) instead of (2017) ?

2020

- on line 573 : Fig. 14 instead of Fig. 16

Corrected

- line 613 : decreases from 87 %?

"from" inserted

- line 693 : Sourdeval instead of Sourdevall

Corrected

- on the legend of Fig. 3 : The acronyms SF and SC could be detailed

We have included the details in the caption.

- in Table 3 : on column 9 : some totals are wrong :

Thanks for the tips. We have made the corrections.

94.4 instead of 69.5 on line 1

5.7 instead of 30.5 on line 2

90.9 instead of 71.3 on line 5

9.1 instead of 28.7 on line 6