

EGUSPHERE-2026-1030

## Automated Detection of Low-altitude Isolated Mesospheric Radar Echoes Using YOLOv8: Evidence for a C-Layer Phenomenon near 60 km Altitude?

Yadu Krishnan Krishnakumar, Toralf Renkwitz, and Andreas Ahrens

### Answers to Reviewer #1

We are grateful for the detailed and constructive review of our manuscript. Please find below our answers to the remaining points.

- Line 100. What is the mean average precision index, and what do the numbers following mAP mean?

Thank you for this comment. We agree that these evaluation metrics require clarification for readers outside the machine learning domain. We will add a brief explanation to the manuscript at this point.

Mean Average Precision (mAP) is a standard evaluation metric in object detection. Precision measures what fraction of the model's detections are correct, while recall measures what fraction of all true targets the model correctly found. mAP summarises the trade-off between these two quantities into a single score, where higher values indicate better performance. The numerical suffix denotes the Intersection over Union (IoU) threshold, the minimum spatial overlap required between a predicted and a ground-truth bounding box for a detection to be counted as correct. mAP50 uses a single threshold of 0.50, while mAP50-95 averages performance across thresholds from 0.50 to 0.95.

- Line 101. What is an "anchor"? Why is it used?

We agree that the term "anchor" requires clarification for readers unfamiliar with object detection methods. We have revised the manuscript text as follows:

In earlier object detection models, anchor boxes are predefined rectangular templates of fixed sizes and aspect ratios placed across the image. The model uses these templates as starting references, adjusting their position and dimensions to fit detected objects. While effective, this approach requires careful manual tuning of anchor dimensions to match the expected sizes of target objects. This can limit flexibility across different datasets.

- Line 107. Define InSAR.

We will include the definition of InSAR in the revised manuscript as follows:

InSAR (Interferometric Synthetic Aperture Radar) is a remote sensing technique that uses radar signals to measure ground surface deformation and topography with high spatial resolution.

- Line 109. “detection accuracy of 96.76% mAP50”. What does this mean?

As defined in our response to the comment on Line 100, mAP50 is the mean Average Precision evaluated at an Intersection over Union threshold of 0.50. Here, an mAP50 value of 96.76% indicates that the model correctly detected 96.76% of the true landslide targets in the dataset at an IoU threshold of 0.50.

- Line 113. What are “convolutional layers”?

We will include a brief explanation of convolutional layers to the manuscript follows:

The Backbone helps in feature extraction. It has a series of convolutional layers, which are mathematical operations that scan the image using small filters to detect local patterns such as edges, textures, and shapes. These layers are organised into blocks that progressively downsample the input image while extracting increasingly complex features.

- Line 154, “The model was trained for 50 epochs”. What is meant by an epoch?

The term "epoch" is specific to machine learning and may not be familiar to readers from the atmospheric science community. In the context of neural network training, one epoch refers to a single complete pass of the entire training dataset through the model. The model was trained for 50 epochs, with validation performed after each epoch to monitor model convergence. We will include this clarification in the revised manuscript.

- Table 1, what do “optimizer” and “learning rate” mean?

We agree that these terms should be clarified for readers who may not be familiar with machine-learning terminology. We will revise the manuscript to explicitly define both parameters. The optimizer refers to the algorithm used to update the model weights during training based on computed gradients. The learning rate defines the step size at which the optimizer updates the model parameters during each iteration.

- Line 158, define “confusion matrix”

The confusion matrix, as referenced in the manuscript, is a standard evaluation tool in machine learning that summarises model performance by tabulating the number of true positives, false positives, false negatives, and true negatives, from which key evaluation

metrics such as precision and recall are derived. We will include a brief definition of this term in the revised manuscript.

- Line 209: “Training stabilised after approximately 30 epochs.” What metric is used to determine this stabilisation. The slope does not appear to change from epoch 10 onwards so I’m not sure how stabilisation can be deduced?

We acknowledge that the wording “stabilised after 30 epochs” could be misleading. The training behaviour was assessed qualitatively based on the evolution of the box loss and classification loss shown Fig.5. Both losses decrease rapidly during the first 10 epochs, followed by progressively smaller improvements in later epochs. The smooth downward trend without significant oscillations suggests stable training behaviour.

- Line 230: “representing the approximate boundary of daylight conditions”. I suggest you add “at the altitudes of interest” to this line.

We will incorporate this suggestion into the revised manuscript and add the phrase "at the altitudes of interest" accordingly.

- Line 230, “The vast majority of detections”. You later quote the number of detections occurring due SZAs above 95 deg as 99%, so I suggest you add “(99%)” after “detections” so the reader does not have to get further into the paper to find out this number.

This is a valid suggestion. We will add “(99%)” after “detections” at Line 230 in the revised manuscript.

- Line 336, “prioritisation of detection reliability over completeness”. What do the authors mean by “completeness”. Do they mean visually checking all the spectra for LIME?

Thank you for the clarification request. By “completeness” we refer to recall, i.e. the proportion of true target signals (LIME signatures) that are correctly detected by the model. In this context, prioritising detection reliability over completeness means that the model favours reducing false positives (higher precision) at the expense of missing some true detections (lower recall). This does not refer to manual visual checking of spectra. We will revise the manuscript text to clarify this terminology.

### **Suggestions for improved readability**

We sincerely thank the reviewer for the thorough lingual review and proposed corrections. We will follow all of the suggested changes.