Response to Reviewer 2

We appreciate Reviewer 2 for your constructive comments and detailed suggestions. In this revision, we have taken into account your comments and made a number of major changes to address your concerns, which are highlighted by our point-by-point response below and also shown in the corresponding track-change version.

This manuscript, "A Deep-learning Framework for Retrieving Tropical Cyclone Intensity and Structure from Gridded Climate Data (TCNN V1.0)," presents the development and training of a convolutional neural network (CNN) model to retrieve tropical cyclone (TC) intensity, minimum central pressure (PMIN), and the radius of maximum wind (RMW) from the MERRA-2 reanalysis dataset. These quantities—intensity, PMIN, and RMW—are inherently inconsistent, challenging to observe directly, and typically poorly resolved in reanalysis products and climate models. The authors aim to develop an enhanced tool capable of retrieving TC information from numerical weather prediction (NWP) models, with potential applications in future climate projections using gridded datasets. The novelty of this work lies in applying CNNs to coarse-resolution data (0.5° × 0.625°) to estimate point-scale TC properties, thereby bypassing the need for satellite imagery or high-resolution dynamic models.

Thank you for the positive evaluation of our work and significance, which indeed touches the core of our work. We hope that this revision could further clarify our main contributions and significance, and it is now satisfactory to you.

Regarding model performance, the CNN demonstrates strong results. It is commendable that the model successfully retrieves VMAX values up to 155 knots. However, alternative configurations—such as different optimizers, loss functions, and filter arrangements—warrant consideration. Readers may also be interested in understanding the model's comparative performance in single-output versus multi-output scenarios. Additionally, I recommend validating the multi output retrieval of PMIN and VMAX against empirically derived VMAX values based on PMIN to assess internal consistency.

Point taken. In this revision, we have included a comparison between the single-output and multiple-output model designs in the revised Figures 2, 3, and 4, which show the PMIN/VMAX/RMW values directly computed from grid points using traditional vortex tracking algorithms. These new results offer deeper insight into the physical interpretation of our results as you suggested. Additionally, we have provided new result of the pressure–wind relationship (Figure 5) to further address your concerns about internal consistency.

Regarding other model settings such as optimizers, loss functions, and the number of filters, we have indeed conducted extensive experiments as part of the hyperparameter tuning process to achieve the best-performing model reported here. Some of these details are too technical and mostly relevant to DL implementation/tuning, and so we have chosen to present only a subset of key sensitivity experiments in Section 3. These include variations in domain size, kernel size, number of filters/layers, data sampling strategies, and seasonal stratification, which we found to be most impactful and physically meaningful as presented in Section 3.2.

The model is trained without differentiation across domains or seasons. It would be valuable to investigate whether dividing the model by domain or seasonal regime could reduce prediction errors. This is especially relevant since TCs exhibit basin-specific structural differences and seasonal variability.

Great suggestion. In fact, our previous Figure 10 could address some of your issues, which shows the better performance of our DL model during the peak season using all data for training. Further dividing the data into sub-season/domain for training would, however, result in a much smaller sample size for training. In fact, when we explored this approach during our early model development, we noticed that there are on average < 1 TC per month during January-April, and so the entire 40 years of training data contains < 100 TC data points, which is hard to train any DL model. As such, we decided to train our model with all data points and then test the model for individual month to increase the model robustness. We have added some further discussion on this in this revision (Section 3.3) to make this point clearer.

Concerning the model domain, the sensitivity analysis helps justify the choice of an 18° × 18° domain size. The experiments span the North Atlantic (NA), Northwestern Pacific (WP), and Northeastern Pacific (EP), where this square domain is assumed to encompass both the inner and outer TC cores. However, it would strengthen the manuscript to explain the domain selection process in more detail. A schematic or graphical representation of domain placement relative to the TC center is highly recommended.

In this revision, we have now added more justification why a domain size of 18x18 is chosen. This is indeed one of the important model sensitivities for which we have included analyses with other domain sizes as presented in Section 3.2. Note that all of these domains are centered on TC center locations as described in Section 2. We have revised both of these sections and related discussions so readers can see how we set up the model domain for our DL model development and applications.

Another point of interest is the model's ability to handle multiple cyclonic systems within a single domain. In active basins such as the NA and WP, it is not uncommon for two or more systems to exist simultaneously. Training the model to distinguish between multiple TCs and to identify respective parameters would be a useful enhancement.

This comment is related to your comment about the domain size right above. In this study, our main focus is on retrieving TC intensity under the assumption that there is a single TC in the domain, which is needed so we can examine whether TC intensity can be retrieved from some given ambient environments. Our justification for this assumption is that a TC can leave some imprint on its surrounding environment that allows DL models to learn and predict TC intensity even at a coarse resolution. Handling several TCs within a domain size larger than 18°x18° will be more complicated, as multiple TCs could exist and leave different signals on ambient environment that DL models cannot learn due to the scarcity of those multiple-TC environments. This is the reasons why we fix our domain choice of 18°x18°, which prevents the inclusion of multiple TCs. This point has been now further explained in the domain selection for clarity per your comment.

The manuscript needs to thoroughly revised for the typos and scientific notations. For example, between lines 190 and 195, alternative to "data points" may be chosen for clarity. Similarly, there is a missing period at the end of the sentence on line 315.

Point taken. We have tried to revise the work as much as we can, and hope that this revision will be better presented now.

Overall, this work represents a valuable contribution that could serve as a classifier for extracting TC information from reanalysis datasets. Looking ahead, the authors may consider extending the model to capture asymmetric TC structures. Since the best-track datasets include quadrant-specific wind radii (R1–R4) starting from 2014, future models could be trained to retrieve quadrant-based features rather than only RMW.

Retrieving more TC information such as rainfall rate, different storm radii, outermost radius, or translational speed is certainly an area that we wish to further pursue and hope to present these efforts in our upcoming study. Once again, we thank Reviewer 2 for your very throughout comments and thoughts, which have improved our work significantly.