1. **-TRIVIAL-** Page 4, Line 43, Right Column- The paper cited [18] doesn't have 7 inputs as looking from the dataset file (from github) and the mauscript itself. It looks to me that there are 5 inputs. Can you recheck this?

   ANSWER: In line 91 in their code:

   ```
   scaler_datafile_1 = scaler1.transform(datafile_1)
   X = scaler_datafile_1[:,range(0,6)]   # input variables
   y = scaler_datafile_1[:,range(6,11)]  # output variables
   ```

   Line 130:

   ```
   input_dim = 6   # no. of input variables columns
   output_dim = no_of_output_nodes #no. of output variables columns
   from collections import OrderedDict
   ```

   Shows that they have used 6 inputs.

   It is a typo in our paper, "7" should be changed to "6". And the we should add "the clad-ref-index".
   CK's comment: We should be much more careful about these. Mistakes like these damage our projection of surety.

   AZ: I am sure that I used the same inputs in our experiments.

   Fix: AZ (double check first and update the paper)

   Fixed

   Reply is written

2. **-TRIVIAL-** Also, have you used the dataset of [18] along with your own data making it finally as 1117+432 data samples. If yes, how did you merge both datasets as proposed manuscript (d1,d2,d3,etc) data sets has different inputs parameters than [18]. What happens to non-common columns like number of rings input feature as mentioned in [18].
   ANSWER: No, we have not.

   Reply will be written Reply is written

3. **-TRIVIAL-** Page 7, Line 43, Right Column- Github link for the code is broken. I hope you will correct it. Looking forward to check your code along with data samples you have used to obtain results in this manuscript.
   ANSWER: I forgot to make the repo public. I will make it

public as soon as we finish with the revision.
CK's comment: Don't do that yet. We will talk about it later.
Reply will be written

4. **-TRIVIAL-** Please, reread the paper to correct some misspelling words. For example, Abstract (line 3): should it be "researchers"? Another example in "Introduction", page 1, column 2, line 35: "Snell's Total" instead of "Snell?s Total".
   ANSWER AZ: To be resolved by HA
   CK's comment: HA will not be able to do this as this is a word to latex issue.

   Fix: I have fixed the ones mentioned, we have to re-read the paper before submission.

5. **-TRIVIAL-** Which kind of performance metric the authors are measuring in Table II? Is it MSE? Please, define it.
   ANSWER: Yes, we should indicate in the table that it is indeed the MSE. CK's comment: Just add the necessary bit.

   It was also mentioned in the text.

   Fixed

6. **-TRIVIAL-** Please, define which is RI (page 4, column 1, line 38).
   ANSWER: TO be answered by AY or HA.

   Fixed

7. **-EASY-** How do the authors know the ANN is not overfitting in Fig. 3? Have the authors employed any validation during training? Please, provide the ANN models' performance on the validation by showing the validation MSE during training.
   ANSWER: Yes, we have employed validation. I have prepared validation curves: with and without augmentation with epochs set to 3000.
   Following CK comment: I believe we should include the plots; He will still ask for the plots even if we choose to just mention it. The plots also show that data augmentation reduced fluctuations on the va set which is a very important result in our work.
   CK's comment: We don't have space for another graph.

AZ: Finally, CK, Please rename the tables files used for the loss curves (should be "our data" not "their data"), also in the corresponding latex code.

Fix: The names should be checked ... **Fixed**

Reply will be written regarding the validation curves if we do not have space.

8. **-EASY-** Please provide the test MSEs for comparison between the proposed method and the ref. [18] (illustrated in Fig. 4).
ANSWER: To be calculated by CK using the data used to plot the graphs.

   Fix: Calculation is done AY will place values on graph... **Fixed**

9. **-EASY-** Page 3, Line 13, Right Column- Why you have specifically used 5 hidden layers? What will be the impact on performance when using less or more hidden layers and neurons?
ANSWER: After several experiments, we noticed that using less than 5 hidden layers will cause the model to under fit, while using more increased the training run-time of the model. Also, using more hidden layers led the model over fit as the number of parameters was very high compared to the number of variables in the feature space.
CK's comment: Sure, we will need to reword this, and others too.

   Fixed **Reply will be written** Reply is written

10. **-EASY-** Page 5, Line 46, Left Column- "This can be linked to the fact that this method uses full-batch learning and therefore under-fits the model[46]"- There is no surity of this reason if the above comment/query in line 43 is logical. So reframe/reconsider this statement.

    ANSWER: We can safely drop this argument, or reframe in a way that we are not absolutely sure as more research is required on this topic. Following CK comment: Will be reframed as to what we have seen. CK: I have removed the sentence as the explanation could have easily extended to 4-5 lines. In the reply we can talk about doing experiments about this. **Fixed**

11. **-EASY-** I suggest that the authors give more information about the stopping criteria of the GAN. As the author pointed out, human judgment has an important role in the generation of images and audio. So, could the known physical laws about PCF and SPR be used to replace the human role in the GAN system to improve the performance? How does the author verify the validity of the augmented samples?
ANSWER: When it comes to physical quantities it is difficult to visually interpret the generated samples. I can not answer the first question. However, this is the mainpoint why we chose the WGAN, because the loss function is interpretable as it measures the distance between the distrubutions of real and generated data in a stable manner. I can provide the loss curve of the WGAN which shows that it is indeed converging and that to a high degree of certainty (probabilistically speaking) that our generated data is valid. In a future research it is possible that we implement a system that validates the generated samples according to the physical laws. CK's comment: We explained this in the paper. We will rephrase it in the reply.
AZ: The loss curve of the wgan is ready.

12. **-EASY-** In GAN and ANN, how many training samples and test samples are used respectively? How are the test samples selected?
ANSWER: For the SPR set: for the 9 folds, we used 336 samples to train the gan and the ann, 48 for validation and 48 for testing. For the PCF set: for the 10 folds, we used about 10% for testing, about 10% for validation and the rest for training. CK's comment: We had a table containing this, we removed it to save space.

13. **-MEDIUM-** Figure 4c, Figure 6- Give more details about the paramaters (d1,d2,d3,refrative index,etc) at which this particular actual blue line graph is obtained.
ANSWER: Following CK comment: The reviewer is asking for the input planes used;

    (a) Figure 4c, analyte 1.34:
    $Re(neff)$: 1.46 $Pitch$ = 0.24μm $d1$ = 0.45μm $d2$ =

$0.75\mu m$ $d3 = 0.35\mu m$

(b) Figure 4c, analyte 1.35:
$Re(neff) = 1.45; Pitch = 0.15um; d1 = 0.25um; d2 = 0.75um; d3 = 0.35um$

(c) Figure 6, analyte 1.34:
$Re(neff) = 1.46; Pitch = 0.24um; d1 = 0.45um; d2 = 0.75um; d3 = 0.35um$

(d) Figure 6, analyte 1.33:
$Re(neff) = 1.45; Pitch = 0.15um; d1 = 0.25um; d2 = 0.55um; d3 = 0.15um$

Fixed

14. **-MEDIUM-** Page 5, Line 43, Left Column- You have stated that [18]'s model has been failing in Figure 4c. It looks to me that the blue line in Figure 4c is for your proposed PCF structure design, while [18]'s PCF design was a simple hexagonal structure with varying rings. Also, there was no analyte (but you have used refractive index 1.35 as shown in this manuscript) in [18]. So if you compare ML models (your's and [18]) only on the criteria of wavelength as stated in Figure 4c, then [18] model is supposed to fail. This comparison doesn't make much sense as both designs and input features are different. If you disagree, then explain this comparison in more detail and mention the design specification values. The similar comparison (your's and [18]) explanation is required for Figure 6. // ANSWER: Following CK comment:
AZ: Even if we say that we changed the number of inputs in [18], our argument is still very weak, because [18]'s model is designed specifically for one dataset. We can definitely compare our model to [18] on their data, but on our data I believe we should just say: We have changed the number of inputs and trained [18]'s on our dataset. Figure 4c indeed proves that their model is specifically designed for one dataset and it is normal and very natural, actually it is supposed to fail on other datasets.

CK's comment: We will have to reword this, it is not supposed to fail, it is not designed to succeed. However, we need to show that the design of the ANN is important for generality. .
AZ: I still have my doubts. The thing is they did not claim generality in their paper, their ANN is designed specifically for that dataset ...

CK: That is our claim. We are saying any random ANN cannot generalize to other planes, ours can. Also, in real life you don't have a dataset to design your ANN for. Optimizing an ANN for a single database is considered cheating as this never happens in reality.

15. **-MEDIUM-** Page 7, Line 45, Left Column- "In our experiments, due to optimized dataset the accuracy of the method proposed in Ref. [18] is also improved."- Following from the previous comments- if the comparison is not for the similar designs then you can't state this improvement. So please elaborate on this.
Following CK comment: A fix added by HA, to be unfixed
Fixed We have to explain the problem in the reply.

16. **-MEDIUM-** Page 7, Line 40, Right Column- Recurrent Neural Networks (RNN) generally rely memory to recognise patterns like in speech recognition or text generation and tells what should/will come after. How do you state/propose here than RNN can be used to discover optimal geometrical shapes?- because all the geometrical shapes might be independent of each other.
ANSWER: I believe we should drop this argument. The reviewer is right, if the shapes are independent then an RNN won't be of advantage in this case.

Following CK comment: The argument will be dropped AZ: The argument is dropped; commented out in the latex file 'paper.tex' line 860

Fixed

17. **-HARD-** How did the authors find the ANN architectures? How long did it take to find the ANN designs suitable for the problem being addressed?

ANSWER: The choice for this regression problem given the type of data we have in the feature space was quite trivial: Multi Layer Perceptron was chosen over other architechtures such as CNN and RNN which might not be suitable for this particular problem. Relying on the results attained in [18] and after efficiently training different MLPs (with different layers, tuning the other parameters ... etc) we were able to find an adequate ANN. As to how long; this is unclear as I do not remember and I believe it should be considered as an offline time.

CK: Not really that hard. We can state we have started using the state-of-the-art as the base and made modifications that are deemed beneficial to this particular domain. We should tell him that the ANN structure may not be the most optimal

Reply will be written

18. **-HARD-** The authors wrote, "The most important contribution of this research is the use of the GAN phase, where the available data is expanded to be used in the training phase". Why use a Generative adversarial network? What would be the advantages and disadvantages of a GAN approach over other possible data generators, such as evolutionary mutation functions? The authors should provide a comparison methodology and discuss differences between GAN and other data generators.
ANSWER: To be investigated in future research.
CK comment: We can talk about this in here and maybe in paper. I am not sure he is requesting experiments. We can avoid experiments by adding it to future works.
AZ: We already mentioned and gave references demonstrating the effectiveness and flexibility of GANs. As they are becoming very common method in data augmentation, we have investigated its power to solve the problem at hand. Furthermore, our study was not comparative in nature with regards to data augmentation methods.
Concerning other methods, more experiments , which are beyond the scope of this study, will be conducted to select the the optimal technique.

19. **-HARD-** In Fig 5, the MSE increase when the augmented sam-

ples exceed 1000. Why does more samples cause performance degradation?
ANSWER: AZ I agree with CK's comment
CK comment: This is simple. Augmentation is not perfect. If we keep adding more augmented data, it dialutes the importance of the real data. Initially this is not a problem as the low number of samples is a larger issue. Therefore, there is a balance to be struck here.

Fixed

20. **-NOT CLEAR-** In Fig 6, in the case of SPR, the calculated loss spectrum has a large deviation under some specific parameters. Could the authors discuss the reasons for the deviation and how to judge the validity of the calculation result? ANSWER: AZ: I could not really understand this question
CK comment: We should explain (in reply) the reason behind this is estimating a new structure rather than interpolating missing values within the same configuration.