We apologize for the incomplete and grammatical manuscript reviewed by the reviewers due to an oversight on my part in uploading the manuscript. We have made our best efforts to respond to all comments individually, as detailed in the following section. We sincerely thank you for your valuable comments on our manuscript.

In the Data Preprocessing section, when deleting the table, we also accidentally deleted a paragraph about the dataset description and data processing. This resulted in reviewers seeing an incomplete manuscript at the time of review and confusion about parts of the manuscript. We apologize for this, and the complete data preprocessing section is shown below, with the previously inadvertently deleted content in red.

The data used in this study are from the GPMDB database[22]. Cheng et al.[16] reorganized the baseline dataset from the GPMDB database, only retained the data of Homo sapiens and Mus musculus species. This work follows the benchmark dataset used in the study of Cheng et al.

The dataset provided by Cheng et al. was not de-redundant. To avoid the model performance deviation caused by the redundancy of the dataset, sequences containing non-standard amino acids and redundant information were removed using the CD-HIT tool[23], where the threshold parameter was set to 0.9. For subsequent references, the two datasets obtained after redundancy removal were named dataset A (Homo sapiens) and dataset B (Mus musculus), respectively. In addition, to ensure that the peptides were not biased in the process of assigning labels according to the method described by Guruceaga et al.[11], a new dataset C (Homo sapiens) was constructed by taking the top 0.025% of the E value ranking in the Homo sapiens data in the GPMDB database[12] as a positive sample and the bottom 0.025% as a negative sample in this study. The length distribution of protein sequences in the above three datasets was briefly counted (Supplementary Figure S1), and the vast majority of sequences in datasets A, B, and C had their lengths distributed in the [10,40) interval, with the maximum sequence lengths of 80, 67 and 94, respectively.

For unequal-length biological sequences, a common method is to pad all sequences to the maximum length in the data set. If this approach is adopted in this study, a lot of noise and useless information will be introduced into the model. It has been shown that most information on protein sequences is retained at the N-terminal and C-terminal ends [19]. To retain the integrity of the sequence information and satisfy the model input, the maximum length is set to 40, for sequences longer than 40, splice the first 20 amino acids at N and C terminal ends to represent the sequence; sequences less than 40, only a small amount of padding is needed to satisfy L=40. The statistical information of data sets A, B, and C are shown in Table 1. The data set is divided into a training set and a test set with an 8:2 ratio.

Reviewers' Comments to Author:

Reviewer: 1

Comments to the Author

1. The authors must check the manuscripts for grammatical mistakes, as there are many grammatical errors. For instance, on page 4, lines 38-39 of the introduction page it should be “For this reason or For these reasons”, and on page 6, lines 11-12 of the introduction page it should have been “alleviate this problem or alleviate these problems”. Similarly, there are many such grammatical mistakes in the entire manuscript.

Thank you very much for your valuable comments on our manuscript, as an oversight in uploading the manuscript led you to review a manuscript without grammar checking. We have corrected the grammatical errors you mentioned in great detail. Again, our sincerest apologies to you for our oversight.

2. On page 4 (lines 12-18), both sentences have the same meaning with just a little difference in the words. It’s advisable to keep only one.

Your suggestion is very correct and pertinent, and we have corrected it.

3. The abstract is the essence of any research article and in this paper, and it is advised to revise the whole abstract. There are many grammatical mistakes in the abstract also. For instance, on page 2 (line 55) “multi-feature representation. Introduces BERT” should be changed to “multi-feature representation by introducing BERT”. On page 3 (line 17), “demonstrate that we proposed model” should be changed to “demonstrate that our proposed model”. And on the same page (3) and in line 22, there is a spelling mistake for Homo sapiens.

Your advice on the abstract section was very valuable and we were very inspired. We have rewritten the abstract section and done a thorough grammar check. Thank you again for your valuable comments.

4. On page 8(lines 58-59) and page 9 (lines 4-5), please provide a clear meaning to the sentence “ To fully capture peptide context, a peptide sequence is treated as a sentence with k amino acids treated as "words".”

We sincerely apologize for any confusion about what we meant by this sentence. We have revised this statement in the manuscript and re-explained the intended meaning in it.

The BERT model was first used for natural language processing tasks, using data in the language we use to communicate, such as English or French. When we introduced the BERT model to solve the protein sequence prediction problem, we consider a protein sequence, e.g., "ACDFERLL" as a sentence, and k consecutive amino acids as a word, e.g., if k=2, then AC, CD, DF, etc. are considered as the words that make up the sentence.

If you have any questions about this explanation, please feel free to comment again. Thank you again for your valuable comments.

5. In the whole manuscript, the expansion for the abbreviation “CIBAM” is not given.

We sincerely apologize for this mistake. Please allow us to sincerely apologize for the confusion caused by the unchecked manuscript we submitted. The full name of CBAM is Convolutional Block Attention Module.

6. On page 9 (lines 59-60) and page 10 (lines 4-5), “The channel attention and the spatial attention mechanism computational procedure as (more details see Supplementary CBAM):” The sentence doesn’t seem complete.

Due to the limitation of space, we put the detailed calculation process of channel attention and spatial attention in CBAM in the supporting material. Perhaps our poor presentation made you think this paragraph is incomplete, for which I apologize profusely.

7. In “Results and discussion”, Page 12, the authors should provide the reason for choosing 40 amino acid sequences for performance comparison.

We erroneously removed a paragraph from the data pre-processing section when we submitted the manuscript. The reason why L=40 was chosen is in the deleted paragraph, and we apologize for this. Also, the experiment in Table 2 explains why we chose the length of 40. Again, we apologize for the confusion we caused when you read the manuscript because of our mistake.

8. Authors should also correlate their study with (i) 10.1021/jacs.2c03858 and (ii) 10.1016/j.heliyon.2022.e12283 and cite them appropriately as both relates to news classes of peptide design and search mechanism from large databases.

Thank you very much for the two reference articles you have given us. We have read them carefully and cited them in the manuscript. Thank you for providing this valuable comment.

9. In figure S1, the author should add representation for X and Y axis for the charts.

Thank you for your pertinent comment on the irregularities of our Figure S1, we have redrawn Figure S1 and added descriptions for the X and Y axis. X-axis is the distribution interval of peptide sequence length, Y-axis is the number of peptide sequences.

10. On page 32, table 2, only the performance of one processing method is discussed as opposed to what is mentioned in the legend of the table.

For model construction, the first thing we need to determine is the length that should be used for the peptide sequences, and most of the treatments are to populate all sequences to the length of the longest sequence in the dataset. The peptide sequence length in dataset A (the main dataset used) is distributed as (0, 80], but more than half of the peptides are less than 40. Therefore, we set up several sets of experiments in Table 2 to determine the most suitable peptide sequence length. It can be seen that when L=40, the ACC of the model is better than other sets of experiments and the computational overhead is moderate, so the peptide sequence length is set to 40 in all subsequent experiments.

11. On page 34, table 4 describes the Performance comparison of different deep learning models but doesn't mention on which datasets the comparison was done.

Thank you very much for pointing out the problems with our manuscript. In writing the manuscript, we omitted to indicate on which dataset the ablation experiments of the model structure (Table 4 results) were performed. Our model structure ablation experiments were performed on dataset A, and dataset A is the dataset we primarily used. Thank you again for providing your comments.

12. Figure 3 (page 39) and figure 5 (page 41) can be kept in one common figure.

Thank you very much for the suggestions you provided. The indicator results on dataset A appear in different sections of the manuscript than the indicator results on dataset B. If combined in one figure, it may not be very convenient for reading and paper layout

13. According to the reviewer's suggestion, the author should also discuss the server's limitations (if there are any) and also the future prospects.

Thank you very much for the suggestion provided by the reviewer. We have open-sourced our model and weighting files on GitHub : https://github.com/leonern/DeepPD. and have very detailed instructions on how to use them. We have also built a user-friendly web server at https://huggingface.co/spaces/xiaoleon/DeepPD-hf.

Reviewer: 2

Comments to the Author

In this paper, the authors present a deep learning method for predicting peptide detectability. The materials and methods section is well-organized. However, since this paper is being submitted for a problem-solving protocol, it would be beneficial for the authors to provide a detailed explanation of the problem and a thorough discussion of the strengths and weaknesses of existing methods.

Thank you very much for your valuable comments on our manuscript. We have reworked the introduction section to provide a more detailed overview of the problems with existing methods and to discuss them more deeply. Previously, we sorted out the existing methodology in the second paragraph of the introduction and discussed the existing methodology in less detail in the third paragraph. We apologize for the poor handling of this part of the work and have taken your comments on board.

Minor comments:

1) The authors used five-fold cross-validation instead of ten-fold cross-validation. It would be helpful if they explained their rationale for this choice.

The main reason is to align with previous work, and we have also performed ten-fold cross-validation, and the performance of the model obtained is essentially the same as the five-fold cross-validation. Thank you very much for your valuable suggestion, and we have added a note in the manuscript about the reason for choosing the five-fold

cross-validation.
2) It would be useful if the authors explained how class imbalance affected the model's performance.

Thank you very much for your advice. However, the datasets covered in this manuscript are category balanced. Previous studies have all been conducted on datasets A and B, and have exhibited unbalanced classification for positive and negative samples (SN and SP differ significantly). We created a new dataset C to explore whether the model's unbalanced classification of positive and negative samples is caused by poor labeling information in the dataset itself.
3) It would be helpful if the authors provided information on the ratio of data used for training versus testing

When we removed Table 1, we accidentally removed the part about dataset partitioning and dataset introduction. This is why the data preprocessing section, looks incomplete. Again, we sincerely apologize for our mistake.

Reviewer: 3

Comments to the Author
The author combines three peptide representation methods and connects three machine learning models before the fully connected layer for classification to build a cutting-edge model. Overall, the logic of the paper should be enhanced (introduction and method section). Terms should be consistent throughout the whole manuscript. The reason analysis of the model performance can be more thoughtful.

Thank you very much for your valuable suggestions on our manuscript. Due to an error on our part, we uploaded a manuscript that was not grammatically checked and was partially missing (when removing Table 1, we accidentally deleted part of the data preprocessing section by mistake). We sincerely apologize for this. We have reinforced the logic of the introductory section and the experimental section on model ablation, and your valuable suggestions have been very helpful.

1.Page 4, line 7-23. The illustration of the term “peptide detectability” is confusing. I suggest the author to rephrase these sentences.

Your opinion is very correct and useful, and we are very grateful for it. The definition of peptide detectability should be stated in only one sentence, and due to our mistake, the sentence with a similar meaning was not deleted. Again, we sincerely apologize for our mistake.

2.The literature review in the second paragraph of Introduction section should be summarized and integrated logically instead of just enumerating them.

With your advice, we realized that our writing about the literature review section was very poor. In the second paragraph of the introduction we have only listed the existing research methods and in the third paragraph we have only made a lesser discussion, which is very weak to illustrate the shortcomings of the previous work and the need for our

work. We have taken your comments seriously and have carefully rewritten the second and third paragraphs of the introduction. Thank you again for your valuable comments.

3.I do recommend a major revision of the introduction first. I do not see a clear objective of the study from the introduction section, and what makes the study differentiated from previous studies.

Thank you very much for your pertinent and very useful comments. We are acutely aware that the summary and discussion of the shortcomings and strengths of previous studies in our manuscript are very inadequate. We have also failed to show the differences in our proposed approach in the introduction section. We have conducted a very deep self-reflection on the above issues and have thoroughly rewritten the introduction section. We would like to thank you again for your valuable suggestions.

4.Please add a section to describe your dataset information.

We apologize for the confusion caused by the data pre-processing section where we mistakenly removed the data set description and data pre-processing. Following your suggestion, we have split the data set description and data preprocessing into two chapters.

5.The last sentence of Data preprocessing subsection is not a complete sentence.

We apologize for the confusion caused by the data pre-processing section where we mistakenly removed the data set description and data pre-processing. Following your suggestion, we have split the data set description and data preprocessing into two chapters.

6.The author stated that “equal-length sequences are obtained according to the methods mentioned in the data preprocessing section”. I do not find it is mentioned above.

When submitting the manuscript, it was requested that the graphs in the manuscript should be in a separate word. When we deleted Table 1, we accidentally deleted the part about dataset description and data preprocessing, which caused you a lot of confusion, and we are sorry for that. The determination of the peptide sequence length is exactly in the section that was deleted by mistake.

7.In Figure 1, the word embeddings module is not displayed. Please specify how you implement the word embeddings. You use a similar way, such as FastText or other approaches to train a model for the peptide embeddings? it is not clear.
From your Figure 1, it seems the L in One-hot encoding section is a fix values 40 ? if it is, you should state that.

In the experimental part, the first thing we need to determine is the length of the peptide sequence, so the first experiment is about it. We determined that the model can obtain optimal performance when L=40 and the computational time is not good to increase too much, and all subsequent experiments set the peptide sequence length to 40. the specific sequence clipping method, which was accidentally removed partly when we removed Table 1, was also removed. We added it back to the manuscript and put the

missing content at the beginning of the word document when we responded to the reviewers' comments.

The specific word embedding is implemented through the nn.embedding() method in the Pytorch framework. nn.embedding() layer does not use algorithms such as word2vec or Fasttext for word embedding, but rather a linear lookup table to obtain a representation vector for each amino acid in the dictionary. The embedding vector for each amino acid is represented by a learnable one-dimensional discrete vector, which is also optimized during the backward gradient update when the model is training. A detailed explanation can be found at https://pytorch.org/docs/stable/generated/torch.nn.Embedding.html

8.Figure 1 cannot be very well consistent with the description in the manuscript.

We have not gone into enough detail in describing each module of the model, which may have caused you some confusion. We have provided a more detailed explanation of each module of the model. We can assure you that in writing the manuscript we have followed the algorithm. we thank you again for the comments you have provided.

9.Transformer and BERT section

It is quite confusing, that the author states “only one Transformer encoding layer is used for attention weighting of …”. It is a self-attention layer displayed in Figure 1. I can not understand clearly why it is a Transformer layer.

How do you combine the features from one-hot encoding, BERT, and word embeddings? The explanation is not clear.

The Transformer and BERT section only tries to explain the mechanism of transformer, while it is not the main point of the study. Since you might only call the model architecture, it is not necessary to spend too much time here, but your feature space construction.

Thank you very much for pointing out the problem in our manuscript, Figure 1, due to our oversight of uploading a previous model image (we had tried the self-attention mechanism before). In the actual model, we used a transformers encoder layer to fuse the word embedding features and the hot encoding features, this is because the multi-headed attention mechanism in the transformers encoder layer can capture the key features of both encoding features very well, we combined the hot encoding features of each amino acid (1* 20) and word embedding features (1*108) of each amino acid into a feature vector (1*128), and the dot product operation inside the multi-headed attention mechanism can well capture the key information present in the feature vector (1*128). This is the reason why we use the attention mechanism here to fuse thermally encoded features and word embedding features.

We have enhanced the Transformer section with instructions on how to use the transformer encoder layer to fuse thermally encoded features with word embedding features. Your pertinent and useful suggestions are greatly appreciated.

10.CNN module, CBAM, and Bi-GRU module should provide some key parameters either in the manuscript or in Figure 1. It would make the paper much more readable.

Thank you very much for this suggestion, we have added a table in the model implementation details section to show the key parameters of CNN, Bi-GRU, and other networks. We believe that the readability and comprehensibility of our manuscript will be improved by accepting your suggestion.

11.Ablation experiments

How do you fuse the three different features? just concatenate them or ? please specify in methods. Which model you used in the feature encoding ablation experiments ? Same question, which feature encoding methods you used in model structure ablation experiment?

We apologize for the confusion caused by the lack of detail in some of the details we wrote in the manuscript. Please allow me to explain your confusion. The model we propose uses three main types of features (thermal encoding, word embedding, and contextual encoding features provided by BERT). The features obtained from thermal encoding and word embedding are fused with a Transformer encoder layer and fed into the feature extraction network A (including 2DCNN, CBAM, Bi-GRU) for feature extraction, while the contextual encoding provided by BERT is fed through the feature extraction network B (including only Bi-GRU) for feature extraction. The higher-order discriminative features obtained from the two feature extraction networks (A and B), in which we did not do any fusion, were simply spliced together and fed into the fully connected layer for classification.

Allow us to rephrase the logic of the ablation experiments in the manuscript. First, we needed to determine the length of the peptide sequence to be used, hence the experiments in Table 2. Secondly, it needs to be determined that all three features we use are useful for the model to predict peptide detectability, hence the feature ablation experiments in Table 3, which were performed on the model shown in Figure 1 (DeepPD), specifically, if Feature A1 (word embedding) is not used, then the only features that enter the Transformer encoder layer are the hot encoded features (L*20, L=40), and similarly for FeatureA2; if Feature B (contextual features) is not used, then BERT and the corresponding feature extraction network B (containing only Bi-GRU) is removed completely. Finally, we need to show that each of the network structures we use is useful for model prediction, so we need to perform ablation experiments on the model structures, and here we use three features (unique thermal encoding, word embedding, and contextual information). Since the contextual information provided by Bert and the corresponding feature extraction network B (which contains only Bi-GRU) cannot be used for the model structure ablation experiments (if the contextual information is removed, the corresponding feature extraction network B has to be removed, which duplicates the feature ablation experiments above). Therefore, we will only discuss the model structure ablation experiments for Feature A1 and Feature A2 feature extraction network A, the results in Table 4.

Once again, we apologize for any confusion our lack of clarity may have caused you when reading the manuscript.

12.Have you repeated your experiments more times and checked the average results

since the feature combination involved with Feature B are all very close each other? A standard deviation should be provided. Table 4 has the same problem.

Yes, for each experiment we did a 10-fold cross-validation and a 5-fold cross-validation (the previous study mainly used the 5-fold cross-validation, so the results shown in the manuscript are the 5-fold cross-validation results), strictly speaking, there is some fluctuation in the results of each experiment, but overall there is an improvement in the model results after using Feature B (each metric improvement (all within 1%). In both Tables 3 and 4, we only show the best results for different models. We have taken your advice and added the results of the five-fold cross-validation to the supplementary material.

13.Page 15 line 17. Please give more evidence to support that DeepPD has a richer feature space with less redundancy

If Feature A2 is added to Feature A1, and if the features provided by Feature A2 and Feature A1 are identical, or if Feature A2 is a redundant feature, then the performance of the model will not be improved, and will even be degraded due to the redundancy of features. Another phenomenon of feature redundancy: when the model is trained, it will be overfitted very quickly (which can be judged by observing the loss of the validation set during model training, which we did not see on DeepPD during training), resulting in a reduction in the generalization ability of the model (because the model accepts redundant features and is overfitted to the training data). When we performed the feature ablation experiments, no overfitting was observed during training and our proposed model also performed optimally in the final model generalization evaluation experiments. Based on your suggestion, we will add a comment to the sentence you mentioned. Thank you very much for your valuable suggestions, which are greatly appreciated.

14.Page 16 line 6-11. Please make the terms consistent between the manuscript context and Figures.
What are the “original features, the high-dimensional abstract feature after feature extraction, …” represent in Figure 4?

Thank you very much for your advice. We have rechecked the terms in the manuscript and made them consistent in the context and the diagrams.

As submissions cannot include diagrams (they have to be uploaded separately), we have removed the diagram names and diagram descriptions from the manuscript as well, which makes some of the images difficult to understand, for which we apologize. We have re-instated the diagrams in the manuscript.

The original features in Figure 4 refer to Features A1, A2, and Feature B, while the higher-order abstract features are those extracted by the corresponding feature extraction network. We apologize for any confusion in your understanding of the manuscript due to our mistake. We will add more detailed descriptions to the images again.

15.Page 16 line 22. How do you fuse of two kinds of peptide information? Treat them parallelly and concatenate them together. or you have other strategies to fuse them effectively. From my perspective, a simple combination might cause poor performance.

One solution might using a transformer to fuse two source features. You can refer to this paper:
Zhang, Z., Xu, M., Chenthamarakshan, V., Lozano, A., Das, P., & Tang, J. (2023). Enhancing Protein Language Models with Structure-based Encoder and Pre-training. arXiv preprint arXiv:2303.06275.

Thank you very much for the references. We have simply spliced the higher-order features obtained by extracting Feature A1, A2, and Feature B by the corresponding feature extraction networks respectively. We did not take steps to fuse the two because the higher-order features contain different key features that contribute to peptide detectability prediction, and we only wanted to keep as many of the two different features extracted for model prediction as possible. Simple splicing only would also allow for some degree of performance improvement in our proposed model. Thank you again for the references and suggestions.

16.Page 17 line 20-22. The author states the reason might be the two datasets share many common features, while the model has the same architecture, and they do have many common features. I think more explanation should be presented here.

Thank you very much for your valuable advice. In the chapter on model generalization and transfer learning ability evaluation, the model showed better results in transfer learning. We think the reasonable explanation is that the peptide sequences in the two data sets contain many "common features". Paper：

Cheng H, Rao B, Liu L et al PepFormer: End-to-End transformer-based siamese network to predict and enhance peptide detectability based on sequence only, Analytical chemistry 2021; 93:6481-6490. provides the same explanation.

To ensure that there is only one variable (dataset) in the experimental process, we need to ensure that the same model is used. The model has good transfer learning ability, which can only be interpreted from two aspects: 1. There are many common features in the two data sets (A and B) used for transfer learning ability evaluation, which makes the model train on data set A (B), but test on data set B (A) still has good performance; 2. Our proposed model indeed has strong feature extraction capabilities, and our proposed feature representation scheme can effectively characterize peptide sequences. The first point is the main reason, if two datasets do not have "common features", the results of transfer learning will be bad even if the model performance is good. If two datasets have more "common features", but the model's feature representation and feature extraction ability are poor, the results of transfer learning will also be bad. I hope our answers can eliminate some of your doubts, and thank you again for your valuable and useful comments.

17.Page 17 line 43. Where is Table 7?

Table 7 shows the test results of the different models (DeepPD and the four existing models) on dataset C. If possible, you can double-check if Table 7 is available in the manuscript provided by the publisher; if it is not, then we have made a mistake in uploading the table and we sincerely apologize for this.

18.Page 17 line 43. The author state the previous model performance in dataset C is poor and imbalanced (please cite them), and then the five methods in this study is more balanced. However, four of the five methods are all previous studies. It is self-contradictory.

We are also confused about your question and wonder if it is because we did not express it clearly. In the last paragraph of the section ”Model generalization and transfer learning ability evaluation”, we retrained and retested the five models DeepMS, CapsNet, PepFormer, PD-BertEDL, and DeepPD on dataset C. The experimental results show that all five models are relatively balanced for the classification of positive and negative samples on dataset C (the results are in Table 7). On the datasets provided in the previous study (A and B), the five models significantly outperformed the classification of positive samples than negative samples (sn values were about 15-20 percentage points higher than sp values, and the models classified positive samples more correctly). It is known that an unbalanced classification of positive and negative samples can have an impact on model performance (the positive and negative samples in the dataset are balanced, but the model classifies the positive and negative samples are unbalanced). To investigate whether this unbalanced model classification result is caused by the poor labeling information of the datasets (A and B) used in the previous study, we constructed dataset C for comparison experiments.

We apologize for the confusion caused by the unclear presentation in our manuscript.

19.Page 18 line 26. Fine-tuning is used for the downstream prediction task. How do you fine-tuned BERT model for feature extraction/embeddings. The embeddings would be extracted from pre-trained model.

We rechecked the manuscript and there is a misrepresentation regarding the contextual information about the peptide sequences obtained from BERT. We only obtained contextual information from BERT and did not use our data to fine-tune the BERT model, which is a computationally intensive task. Thank you very much for pointing out the problems in our manuscript, which we sincerely appreciate.

20.Page 18 line 54. The author state they reconstruct a new dataset. Is it Dataset C? However, Page 17 line 35, the author state the dataset C is provided by previous studies [16]?

A detailed description of how dataset C was constructed is given in the data preprocessing section (part of it was mistakenly removed when the manuscript was uploaded, we have noted the mistake and filled in the new manuscript). Dataset C is our newly constructed dataset; the datasets used in the previous study were A and B. We apologize for any confusion caused by our mistake.

21.A webserver is needed for the wide usage of the problem solving protocol.

We have open-sourced the model and weights to the GitHub repository with detailed instructions for use. We have also built a user-friendly web server at https://huggingface.co/spaces/xiaoleon/DeepPD-hf.Thank you very much for your valuable suggestion.

22.Grammar errors needed to be double-checked:
Page 2, line 55-59; Page 4 Line 12-30 (no conjunction words); Page 5 line 30; Page 6 line 14-17; Page 6 Line 47; Page 13 line 56; Page 18 line 49;

Thank you very much for pointing out the grammatical problems in our manuscript. After receiving your comments, we have redone the grammar check of the manuscript and rewritten the sentences you have marked.

23.The paper should be polished regarding English writing. Some sentences are redundancy and can be written in a clear and concise way (Page 14, line 3-9; Page 17 line 1-12)

Thank you very much for this valuable comment. We have rechecked the manuscript for grammar and have carefully rewritten those sentences you pointed out. Thank you again for pointing out the grammatical and syntactical errors in our manuscript.

24.A few more suggestions: please spell out the full name of the abbreviation in the manuscript, such as MS, CNN, CBAM, etc.

Thank you very much for pointing out the shortcomings in our manuscript. We did not check in detail whether the full English full name was given for the first occurrence of the abbreviation. Thank you very much for your suggestion, we have rechecked all the abbreviations that appear and given the full English spelling when they first appear.