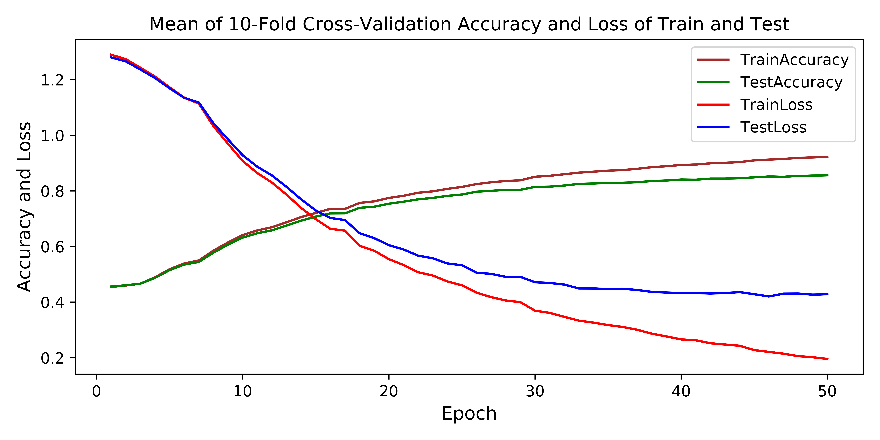
Thank you for your comment. Based on your comments, we have completely revised this paper. Those grammatical problems, spelling problems, parameter errors and many unreasonable problems have been thoroughly corrected. At the same time, we adjusted the experimental procedure to cause slight fluctuation of the prediction results.  
**Major point:**

**1. The Results section is too short. There are only ~40 lines in the current version. The authors should expand and rewrite the Results section thoroughly.**

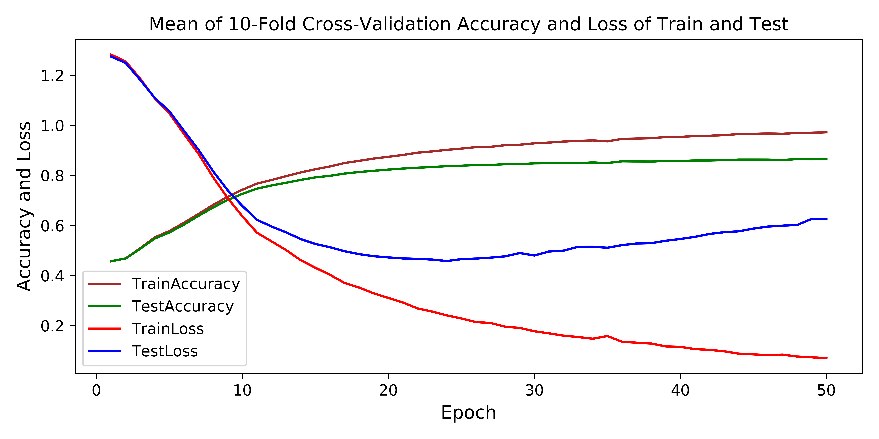
Thank you for your comment. Based on this comment, we have made the following corrections to the Results section. (1) We moved the first three paragraphs of the original Results section to the Method section and named it "Performance Measurement". (2) We rewritten and expanded the content of the results section. The way of comparing prediction results between different methods is elaborated. We added the visual comparison based on the original comparison, which made the content of the section more substantial and shown some advantages of our method in more aspects.  **2. Currently, the method lacks a well-documented executable standalone software package or an online web-server. Although the original data and codes are available on GitHub, there lacks instructions of how to run or use the codes step by step. A working standalone or online version would benefit users greatly who are interested in using this new method. It will also increase the impact of this work.**

Thank you for your comment. Based on this comment, we have added annotations to the key parts of the program and written instructions for the use of the program to make it easier for interested readers to use our method. At the same time, we uploaded the program to GitHub and added the URL of the program to the abstract. It is convenient for readers to download and use. Since we store the intermediate results and the recorded models (CU model) in the program, the program takes up a lot of space. Please be patient to download. **3. The authors could consider collecting more RNA secondary structure dataset and include them in training and testing. It is known that a relatively large dataset is required to train a deep neural network. Also, from Figure 6, the models might suffer from over-fitting problem to some extent as the training loss continues to decrease but the validation loss tends to increase after ~50 epochs. Maybe having a larger training set could reduce this problem. The authors could also consider decreasing the learning rate or increasing the batch size to see if it helps.**

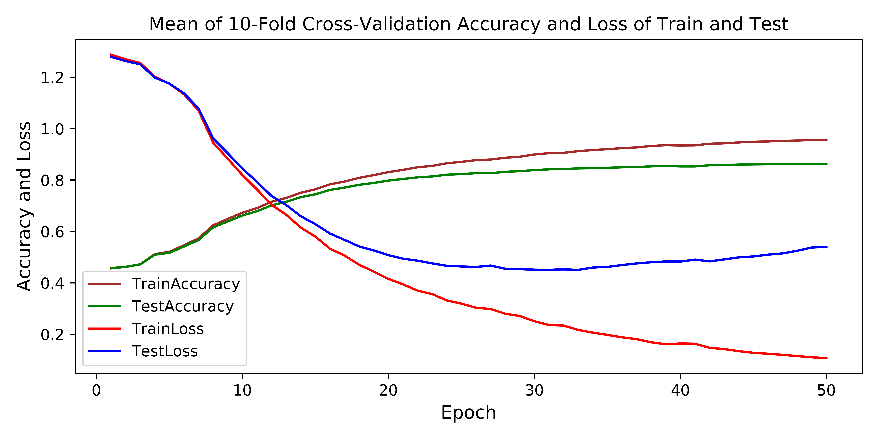
Thank you for your comment. Based on this comment, we have done the following experiments. (1) In addition to the dataset in the paper, we also collected three datasets: tRNA [1], tmRNA [2], and RNA Strand [3]. All data of 5sRNA, tRNA, tmRNA and RNaseP was extracted from the above four datasets. Unfortunately, since the data collected by multiple databases is the same, after data pre-processing the clean data not have increased. (2) Since the dataset cannot be improved, we use your comment to reduce the learning rate or improve the batch to see if it can reduce the degree of overfitting while ensuring the correct rate. We performed three experiments separately: reducing the learning rate alone (reducing the learning rate from 0.002 to 0.001), increasing the batch size alone (increasing the size of the batch from 16 to 32). both reducing the learning rate and increasing the size of the batch (reducing the learning rate from 0.002 to 0.001 and increasing the batch size from 16 to 32). **Figure 1** shows the average loss and accuracy at each epoch in the 10-fold cross-validation experiments of only reducing the learning rate. **Figure 2** shows the average loss and accuracy at each epoch in the 10-fold cross-validation experiments of only increasing the batch size. **Figure 3** shows the average loss and accuracy at each epoch in the 10-fold cross-validation experiments of both reducing the learning rate and increasing the batch size. **Figure 4** shows the average loss and accuracy at each epoch in the 10-fold cross-validation experiments of the original experiment. **Figure 5** shows the average testing accuracy and loss of four experiments in each epoch. It can be seen from those five figures that increasing the batch size alone or both reducing the learning rate and increasing the batch size, not only increases the degree of overfitting but also reduces the prediction accuracy. Therefore, these two ways are not desirable. Although reducing the learning rate alone can slow the overfitting, when training it 100 epochs (**Figure 6**), there was also an over-fitting phenomenon. It can be seen from Figure 5 that the loss of the only reducing learning rate and the original parameters is similar near 50th epoch. However, the correct rate of the original parameters is indeed greater than the case of reducing the learning rate alone. And although there is a slight over-fitting in the training process of the original parameters, it does not affect the prediction accuracy. This may be because during the classification process we will select the node with the largest value in the output layer as the classification result. Although the test loss is increasing, the node of the maximum value has not changed. Therefore, it does not affect the prediction accuracy rate. Comprehensive analysis, using the original parameters (learning rata is 0.002, batch is 16, epoch is 50) best fit the needs of the model.



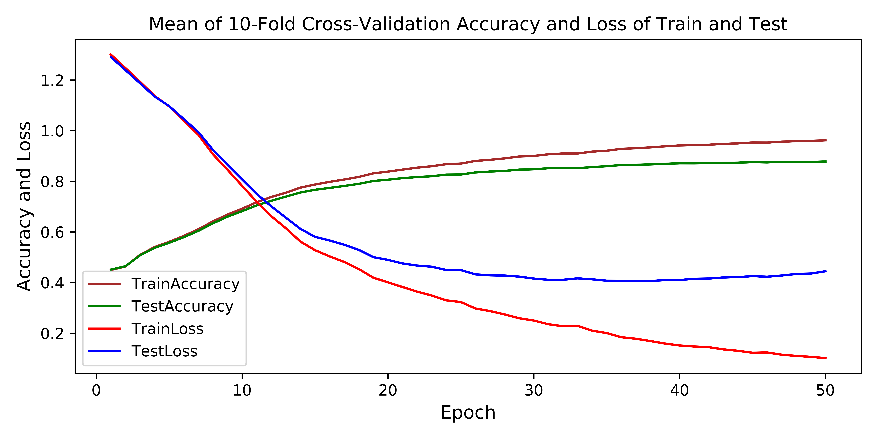
**Figure 1** The learning rate is 0.001 and the batch is 16.



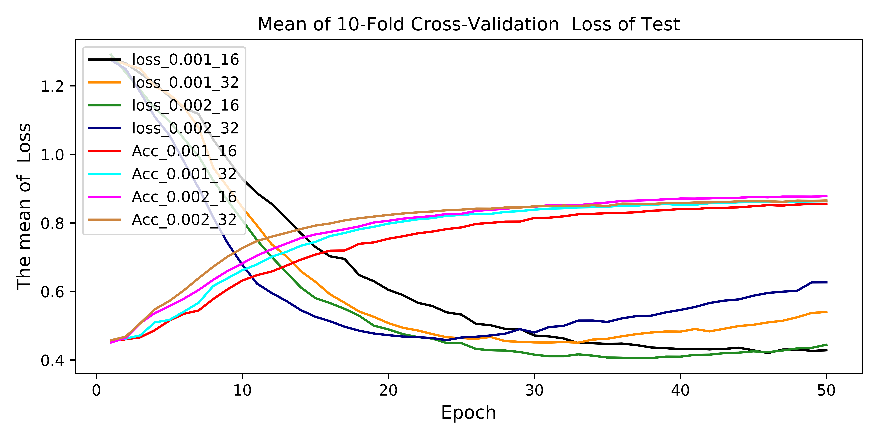
**Figure 2** The learning rate is the 0.002 and the batch is 32.



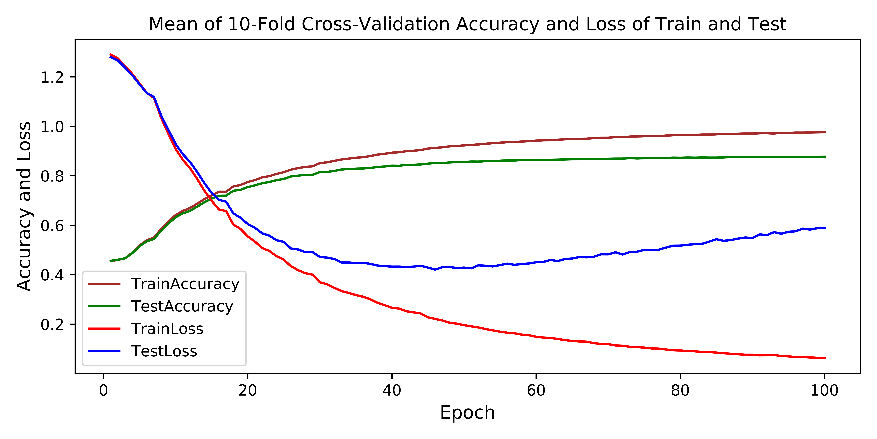
**Figure 3** The learning rate is the 0.001 and the batch is 32.



**Figure 4** The learning rate is the 0.002 and the batch is 16.



**Figure 5** The test loss and accuracy of three correct experiment and the original experiment



**Figure 6** The learning rate is 0.001 and the batch is 16.

[1] F.Juhling, M. Morl, R. K. Hartmann, M.Sprinzl, P. F. Stadler, and J.Putz, “tRNA db2009: Compilation of RNA sequences and tRNA genes,” Nucleic Acids Res., vol.37,no.Database issue, pp. D159-62, Jan. 2009.

[2] E. S. Andersen, M. A. Rosenblad, N. Larsen, J. C. Westergaard, J. Burks, I. K. Wower, J.Wower, J. Gorodkin, T.Samuelsson, and C. Zwieb, “The tmRDB and SRPDB resources,” Nucleic Acids Res., vol.34, no. Database issue, pp. D163-D168, Jan. 1, 2006.

[3] M. Andronescu, V. Bereg, H. H. Hoos, and A Condon, “RNA Strand: The RNA secondary structure and statistical analysis database,” BMC Bioinf., vol 9, p.340,2008

**4. Section 2.1: consider making a table or supplemental table summarizing the data used in the study so that readers do not need to go to the individual data source to check the details of the dataset.**  
Thank you for your comment. Based on this comment, we have made the following corrections to the section of 2.1. (1) We have moved dot-bracket transformation from Introduction to section 2.1. (2) In order to facilitate the reader to understand the experimental data. We added a table to the supplemental materials. Explain the details of original data, experimental data, and clean data.  
**5. Is a single model trained using data from all RNA types or are separate models trained using data from each individual RNA type?**  
Thank you for your comment. We sincerely apologize for having overlooked the issue. Our method is a single model, which simultaneously using multiple types of RNA as learning and training data, to predict the secondary structure of multiple types RNA. We have added an explanation of the problem in the fourth paragraph of the Introduce section.  
**6. All figure captions need more details, information, and explanations. For example, Figure 1 need more explanations of the colors, lines, symbols, etc.**

Thank you for your comment. Based on this comment, we have written more information into their respective captions. That information can help readers better understand those figures.  
**7. Figure 3, 4, and 5 are sub-parts of Figure 2. The authors could consider to merge them into Figure 2 as sub-panels or move them to supplemental materials. Also, for Figure 3, the mathematical expression is not in the correct format, missing proper subscripts and parentheses.**  
Thank you for your comment. Based on this comment, we have moved Figure 3, 4, and 5 to supplemental materials. And we have modified Figure 3 to make that mathematical expression in the correct format.  
**8. The authors could consider expanding Figure 7 and making a figure to illustrate the complete algorithm of IBPMP.**

Thank you for your comment. Based on this comment, we have extended the original Figure 7 to include the IBPMP algorithm. It enriches the content of the figure and makes it easier for readers to understand IBPMP algorithm.  
**9. The authors could consider moving the first 3 paragraphs of the Results section to a sub-section of Methods whose title could be "Performance Measurement".**

Thank you for your comment. Based on this comment, we have moved the first three paragraphs of the original Results section to the Method section and named it "Performance Measurement".  
**10. The English writing needs revision throughout the manuscript to make it more concise and clearer. Also, the use of capital letters may not be necessary in some places.**  
Thank you for your comment. Based on this comment, we have completely revised the paper to make its sentences more concise. At the same time, many irregular capital letters were changed to lowercase.

**Minor points:**  
**1. Line 15: it is better not to have reference in abstract.**

Thank you for your comment. Based on this comment, we have removed the reference in the abstract.  
**2. Abstract could be more concise.**

Thank you for your comment. Based on this comment, we have carefully revised the abstract to make its sentences more concise.  
**3. Line 53-54: consider introducing here the pseudoknots and why pseudoknots are important.**

Thank you for your comment. Based on this comment, we have added the importance of pseudoknots to Introduction. Since the second paragraph of line 53-54 mainly introduces traditional prediction methods, we add the importance of pseudoknots to the beginning of the third paragraph.  
**4. Line 52-66: consider introducing and citing more papers on RNA secondary structure prediction tools.**

Thank you for your comment. Based on this comment, we have cited more RNA structure prediction tools in the original 52-66 line to make the content richer.  
**5. Line 75: Give the full name of IBPMP along with the abbreviation because this is the first time to mention it in the manuscript main body.**

Thank you for your comment. Based on this comment, we have added the full name of IBPMP on the original 75 line.  
**6. Line 83-94: consider briefly summarizing the features and processing steps of DMfold. Move the dot-bracket transformation to section 2.1 Data Collection and Processing.**

Thank you for your comment. Based on this comment, we have moved dot-bracket transformation from Introduction to section 2.1. And described the features and processing steps of DMfold in this part.  
**7. Line 109: naming (correct unit) is inconsistent with Line 200 (correction unit).**

Thank you for your comment. Based on this comment, we have modified all the “correct unit” in the paper to “correction unit”.  
**8. Line 112, 161: change “full connect layer” to “fully connected layer”.**

Thank you for your comment. Based on this comment, we have modified all the “full connect layer” in the paper to “fully connected layer”.