22-Apr-2025  
  
JBHI Ref: JBHI-01395-2025  
Reject/Resubmit (major revision and new external review required)  
  
Dear Dr. Tiangang Zhang,    
  
This letter is to inform you that the peer review process has concluded for manuscript, "A multi-scale neighbor topology guided transformer and Kolmogorov-Arnold network enhanced feature learning model for disease-related circRNA prediction," which you had submitted for possible publication in the IEEE Journal of Biomedical and Health Informatics (J-BHI).  
  
The Associate Editor responsible for your manuscript review has received feedback from independent reviewers and compiled their evaluations. It is the recommendation of the Associate Editor and the Editor-in-Chief that your manuscript requires a MAJOR REVISION before it can be accepted for publication in J-BHI. Please note the J-BHI Editorial Policy that only one major revision is allowed for any submitted manuscript. This means that if the review recommendation for your revised paper triggers another major revision, the paper will be rejected automatically.  
  
Enclosed, please find the comments by the Associate Editor and all the reviewers. I hope that the feedback is helpful for further improving the quality of your manuscript. If you decide to resubmit a revised manuscript, this must be done within 10 weeks (not extendable) from the date of this message. Please quote the above manuscript reference number for all future correspondence.  
  
**\*IMPORTANT: PLEASE RETURN HERE TO SUBMIT ALL FILES:**[https://ieee.atyponrex.com/journal/jbhi-embs](https://ieee.atyponrex.com/journal/jbhi-embs" \t "https://mail.163.com/js6/read/_blank)  
  
Also note that it is mandatory to enter your replies to the reviewers' questions and indicate how you have dealt with their comments in the revised manuscript. Please include your replies in the authors' response section or a separate file with a point-by-point explanation of all the changes made. Please do NOT include it in the cover letter since this is not accessible by the reviewers. For submitting your revised manuscript, please ensure all changes are highlighted in the manuscript to facilitate the review process.  
  
Citing/including papers suggested by the reviewers or the Associate Editor is up to the authors and only if they add value to your work.  Please also report to [fotiadis@uoi.gr](mailto:fotiadis@uoi.gr) any suspicious suggestion by the reviewers.  
  
The authors are responsible to follow the J-BHI publication rules for maximum number of pages, quality of figures, etc. as they are mentioned in: https://www.embs.org/jbhi/prepare-and-submit-your-manuscript/  
On the opposite your article might need to be returned to you and the review process will be delayed and in case of acceptance no publication is possible.  
  
Please carefully read about how to prepare your manuscript and note the mandatory charge for over-length papers as imposed by IEEE in the website of J-BHI (https://www.embs.org/jbhi/prepare-and-submit-your-manuscript/).  
  
Thank you very much for considering JBHI to publish your research work.  
  
Sincerely,  
  
Prof. Dimitrios I. Fotiadis  
Editor-in-Chief  
  
Cc: file  
  
Associate Editor's comments to the authors:  
Associate Editor  
Comments to the Author:  
The reviewers raise many concerns that prevent acceptance of the paper in its current form; however, we encourage you to address these issues and resubmit. While the proposed MKCD model, which integrates a multi-scale neighbor topology-guided transformer and Kolmogorov-Arnold Network (KAN)-enhanced feature learning, presents innovative ideas for circRNA-disease association prediction, there are substantial weaknesses in writing clarity, methodological description, experimental validation, and analysis rigor. Key concerns include insufficient background to contextualize the study’s novelty, lack of detailed explanation and sensitivity analysis regarding the Random Walk with Restart (RWR) strategy, unclear parameter configurations in KAN, limited ablation studies isolating the effects of different model components, and scalability issues for large-scale graphs. Furthermore, the manuscript contains inconsistencies in formatting, needs clearer figure annotations, and would benefit from additional datasets, expanded evaluation metrics, and deeper biological validation of predictions. Overall, although the manuscript demonstrates potential with promising concepts and preliminary results, significant revisions are necessary to strengthen the scientific foundation, experimental design, and clarity of presentation.  
  
For submitting your revised manuscript, please ensure all changes are clearly highlighted in the manuscript and explained in detail in the rebuttal to facilitate the review process.  
  
Reviewers' comments to the authors:  
Reviewer: 1  
  
Comments to the Corresponding Author  
The research propose a multi-scale neighbor topology-guided transformer with Kolmogorov-Arnold network (KAN) enhanced feature learning for circRNA and disease association prediction, termed MKCD. Here are some issues:  
Writing:  
1.The introduction’s background section currently provides a concise overview of circRNA-disease association prediction but lacks sufficient depth to contextualize the study’s novelty.  
Method:  
1.The AMNE module employs Random Walk with Restart (RWR) to generate multi-scale neighbor topology embeddings. How are the parameters configured in RWR? Are they empirically predefined or learned during training? Are distinct parameters values applied to different neighbor scales?  
2.Has experimental validation been performed to assess how variations in parameters in RWR affect the coverage and quality of multi-scale topologies (e.g., balancing local specificity and global context)?  
3.Given that AMNE determines the importance of neighbors at different scales through learnable attention weights, how does the method ensure stability in the learned weights despite the inherent stochastic nature of random walk paths?  
4.The neighbor topologies generated by RWR may be constrained by walk length and restart probability. How do you ensure that the generated topologies cover sufficiently broad neighbor ranges while avoiding the introduction of excessive noise nodes (e.g., irrelevant high-scale neighbors)? Is this achieved by truncating walk lengths or setting node similarity thresholds to filter out noise? If so, have corresponding ablation experiments been conducted to quantify the impact of different noise filtering strategies on model performance?  
5.RWR is one of the common methods for generating multi-scale neighbors, but other approaches (e.g., meta-path-based sampling, graph diffusion algorithms) can also achieve similar goals. Have comparative experiments been conducted to evaluate the performance differences between RWR and other multi-scale topology generation strategies? For instance, does RWR offer distinct advantages compared to alternatives like simple BFS traversal or Personalized PageRank?  
6.The paper mentions that low-scale (one-step) and high-scale (d-step) neighbors contribute differently to node feature learning. How is this "difference in contributions" specifically quantified? Is it measured through metrics such as visit frequency statistics in RWR or path weights between nodes? For instance, does the model inherently assume that one-step neighbors necessarily contribute more than multi-step neighbors?  
7.For large-scale heterogeneous graphs, is the computational complexity of RWR prohibitively high? Are approximation algorithms adopted to improve efficiency? When introducing new node relationships, does the model require retraining?  
8.During the dynamic updating process, do the weights of low-scale and high-scale neighbors change with the training process?  
9.Have experiments been conducted to isolate the effects of low-scale and high-scale neighbors? For instance, when only one-step neighbors or exclusively high-scale neighbors are used, how significant is the performance degradation of the model?  
10.How are the spline functions parameterized in KAN? How are the number, positions, and smoothness constraints of spline knots determined? Are these parameters sensitive to the final performance? Do vanishing/exploding gradient problems exist?  
11.The computational complexity of spline functions is higher than that of traditional activation functions. Does KAN significantly increase model training/inference time?  
12.The paper introduces RWR and KAN strategies. Does the increased parameter count lead to significant growth in computational overhead, thereby limiting the model's ability to scale to large-scale datasets? How is the scalability of the model?  
Formats:  
1.Is the data at the end of the paper necessary?  
  
  
  
Reviewer: 2  
  
Comments to the Corresponding Author  
The authors present a model based on multi-scale neighbor topology guided transformer and KAN enhanced feature learning for predicting the disease-related circRNAs. One key innovation is a dynamic multi-scale neighbor topology-guided transformer is designed and it leverages the multi-scale neighbor topologies to guide the learning of relationship among circRNAs, miRNAs, and diseases. It is also novel that Kolmogorov-Arnold Network (KAN) based strategy is proposed to learn the global and local dependencies within the features of a pair of circRNA and disease. The ablation studies further validate the effectiveness of these innovations. The comparison experiments demonstrate that the proposed method outperforms the compared advanced methods. I have the following concerns regarding the manuscript.  
  
1. In Section II B part, 3rd paragraph. The authors state that the similarity between a pair of circRNA (miRNA) nodes is calculated based on their associated disease sets. Please provide the formula for calculating the circRNA (miRNA) similarities.  
  
2. Formula (4) gives the probability distribution of a random walker reaching all nodes at time t, while Formula (5) defines the adaptive fusion process from scale 0 to scale t. Since two formulas use the same symbol "t", distinct symbols are recommended if their have different meanings.  
  
3. In the 1st paragraph of Section II D part, the authors mention the dynamic guidance and they also reconstruct the multi-scale neighbor topology embedding matrix to obtain the changed topology in the 3rd paragraph. It would be better to clearly state that the embedding matrix is dynamically evolving.  
  
4. Figure 4 illustrates the node feature learning process guided by multi-scale neighbor topology. I recommend explicitly giving the specific meanings of symbols "⊗" and "⨀" within Figure 4 for clarity.  
  
In Section II F part, first paragraph. It is stated that KAN outperforms MLP in capturing the complex relationships between pairwise attributes. Please provide the experimental results obtained by replacing KAN with MLP and include an analysis of the results accordingly.  
  
6. In Section III A part. Regarding the "Adaptive multi-scale neighbor topology embedding construction strategy", the maximum scale is set to 2. Please explain how to select the parameter and list the corresponding experiment results.  
  
  
Reviewer: 3  
  
Comments to the Corresponding Author  
Identifying disease-related circRNAs is crucial for gaining deeper insights into the mechanisms driving disease pathogenesis. This manuscript introduces a novel method, termed MKCD, designed to predict candidate circRNAs associated with diseases. The proposed method contains several novelties: (1) an adaptive multi-scale neighbor topology embedding construction strategy, (2) a dynamic multi-scale neighbor topology-guided transformer, (3) a feature fusion with feature-gated network, and (4) a CNN and KAN joint learning strategy for pairwise feature learning. My comments and suggestions are outlined below.  
  
(a) On page 2, line 58, the manuscript utilized the data about circRNAs and diseases (circRNA-disease associations, circRNA similarities, and disease similarities) for prediction of circRNA-disease associations. Moreover, it also integrated miRNA-related data (circRNA-miRNA interactions, disease-miRNA associations, and miRNA similarities). It is suggested to explain why the miRNA-related data is also utilized.  
  
(b) On page 4, line 43, the manuscript designed a multi-scale neighbor topology embedding construction strategy. It is suggested to explain how the scale size was determined.  
  
(c) In formula (4), \theta\_i(0) is initialized as a one-hot vector. Please explain the practical meaning of this kind of initialization.  
  
(d) On page 5 line 41, the manuscript constructed a feature gating network to fuse {\hat{G}}^{(l)} and G^{(l-1)}. If the annotation for {\hat{G}}^{(l)} is provided in Figure 1, it would better help readers understand the content in the main text.  
  
(e) In terms of the ablation studies (Table 1), the manuscript simultaneously removed both DMTT and FGN. Please explain why these two components should be removed together for the ablation experiment, rather than removing them individually to analyze their respective contributions.  
  
  
Reviewer: 4  
  
Comments to the Corresponding Author  
The manuscript entitled "A multi-scale neighbor topology guided transformer and Kolmogorov-Arnold network enhanced feature learning model for disease-related circRNA prediction," which proposes an integrated analysis of circRNA-disease associations using multi-scale neighbor topology and Kolmogorov-Arnold networks, addresses an interesting and important problem in computational biology. Despite presenting innovative ideas and showing promising results, the manuscript requires several revisions to improve its theoretical foundation, experimental validation, and overall clarity.  
Minor revisions needed:  
1.A comparison of computational complexity and runtime efficiency between MKCD and the baseline methods would provide a more comprehensive evaluation.  
2.Multiple grammatical errors and awkward phrasings throughout the manuscript affect readability. I only list some of the problems in the following pages, please check your main  text carefully.  
a.page 2, "Traditional transformer focus solely..." should be "Traditional transformers focus solely..."  
b.page 5, "...to overcome the problem of single-head attention easily falling into local optima..." is awkwardly phrased.  
3.The paper mentions specific parameter values (e.g., restart probability λ=0.7) without justifying these choices or exploring how performance varies with different parameter settings.  
4.The conclusion section would benefit from a more balanced discussion that includes not only the strengths but also the limitations of the proposed model.  
5.In the ACK component, you used a 2-layer KAN network. Please explain why KAN was preferred over other neural network architectures for learning global feature dependencies.  
  
Reviewer: 5  
  
Comments to the Corresponding Author  
This paper presented a novel approach for encoding the relationships among circRNA, miRNA, and disease node features, while effectively learning and integrating both global and local features of node pairs to predict disease-related circRNAs, achieving better performance. The overall structure of the paper is clear and concise. However, there are still some problems that need to be addressed. My detailed comments are as follows:  
  
1. The authors use FGN to realize the fusion of different types of features. Is this approach superior to other feature integration approaches such as concatenation, attention mechanisms, etc.?  
  
2. The authors should carefully check the manuscript and correct formatting inconsistencies, such as paragraph indentation, spacing, and alignment, to ensure compliance with the journal’s guidelines.  
  
3. To better demonstrate the model’s superiority, the authors should consider including additional evaluation metrics. Furthermore, if feasible, validation on additional datasets would help strengthen the claims regarding the model’s generalization capability.  
  
4. The description of the ablation experiments appears inaccurate. The second ablation case simultaneously removes both DMTT and FGN, yet the manuscript describes the modules as being removed sequentially, which could cause confusion. Additionally, the claim that “DMTT contributes the most” may not be fully justified. Based on the results from the second and third ablation cases, while DMTT does impact model performance, the evidence does not clearly demonstrate that its influence is the most significant compared to other components. The authors should clarify the ablation methodology and provide more rigorous analysis to support their conclusion.  
  
5. The authors should further analyze the model’s unverified predictions to evaluate their potential biological significance. This would more convincingly demonstrate the model’s capability to discover novel associations.  
  
  
Reviewer: 6  
  
Comments to the Corresponding Author  
The task of the manuscript is to infer the reliable disease-related candidate circRNAs, and a prediction model was proposed based on multi-scale neighbor topology guided transformer and Kolmogorov-Arnold network enhanced feature learning. Several novel components were presented in the manuscript: adaptive construction of multi-scale neighbor topology embedding, dynamic multi-scale neighbor topology-guided transformer, feature-gated network to estimate the importance of topological and original features, and learning of the global feature relationships based on Kolmogorov-Arnold network. The presented method is compared with several advanced circRNA-disease association prediction models and it gets superior prediction performance. I have several comments for the manuscript.  
  
1) In the first paragraph of the introduction, the authors mention that identifying circRNA-disease associations helps in disease diagnosis and treatment. Please give more details about it.  
  
2) In the section “dynamic multi-scale neighbor topology-guided transformer”, the number of attention heads (h) is set to 4. Could you please explain how h is determined and provide the experimental results?  
  
3) Equation (16) defines the set of learnable functions Ψ^((l)). Please introduce the calculation process of Ψ^((l)).  
  
4) There is interaction between AMNE and DMTT. In the ablation study, when both DMTT and FGN are removed simultaneously, what replacement strategy is performed for AMNE?  
  
5) In the comparative experiments, the authors performed paired wilcoxon tests to show the prediction performance of the proposed method is significantly higher than the compared method. I recommend briefly describing how this statistical experiment was performed.  
  
Reviewer: 7  
  
Comments to the Corresponding Author  
The authors propose a transformer with multi-scale neighbor topology guiding and KAN enhanced feature learning for predicting disease-related circRNAs. The ideas that introducing multi-scale neighbor topology to guide the transformer, and learning the pairwise features by jointed KAN and CNN are interesting and novel. The manuscript is easy to follow and well-organized. Here are my concerns.  
  
-- Page 3, paragraph 4. The disease similarities are calculated based on the directed acyclic graphs. Please provide a brief description of disease similarity calculation process.  
  
-- Page 5, paragraph 3. The attention score matrix  is multiplied by the multi-scale neighbor topology embedding matrix  using the Hadamard product. Please describe the advance of using Hadamard product here.  
  
-- Page 6. Eq. (15) demonstrates the learning process of pairwise global features. What is the relationship between  and  in this equation?  
  
-- Page 7, paragraph 3. For the random walk with restart (RWR) strategy, please analyze the impact of the restart probability  on prediction performance.  
  
-- In order to let the researchers follow the work, please provide the corresponding code.  
  
Reviewer: 8  
  
Comments to the Corresponding Author  
The paper focuses on the circRNA-disease association prediction problem, which is helpful for providing reliable disease-related circRNA candidates for subsequent biological experiments. Multi-scale topology embedding is generated by random walks on a heterogeneous graph. The correlations of the circRNA and disease nodes are encoded by the designed transformer guided by the multi-scale topology. Furthermore, KAN-enhanced learning is introduced to effectively capture the global dependency within the features of circRNA-disease node pairs. The following suggestions are listed to enhance the clarity and quality of the manuscript.  
  
1. Please clarify how the two parameters, s\_β and s\_γ, in formula (20) are initialized, given that they are normalized to obtain β and γ.  
  
2. How is the number of neurons in KAN determined? Please evaluate the impact of varying neuron counts on the model's performance.  
  
3. Section “parameter setting”. The paper listed the number of convolution layers and kernel sizes for the local pairwise feature learning. It is recommended to include the number of channels for each convolutional layer to provide a more comprehensive description.  
  
4. In the ablation study, the authors analyzed the strategies with the highest and second-highest contributions to the model's performance. It is suggested to also analyze the strategy with the least impact, along with a corresponding analysis.  
  
5. The proposed method MKCD is compared with six advanced methods. The paper has briefly described each comparison method. Please summarize the categories of these six methods to help readers better understand which types of methods are being compared.  
  
Reviewer: 9  
  
Comments to the Corresponding Author  
The MKCD model offers an innovative approach to circRNA-disease association prediction, and it proposes multi-scale neighbor topologies, a transformer guided by topology, and KAN-enhanced feature learning. Its circRNA-disease-miRNA heterogeneous graph, coupled with the AMNE strategy, effectively captures multi-scale node relationships. The DMTT module addresses transformer limitations by incorporating dynamic topologies, while FGN and ACK adeptly fuse topological and biological features, capturing global and local dependencies. Experiments show MKCD outperforms state-of-the-art methods, with case studies validating its practical utility. To further improve the work, I suggest considering the following points.  
  
1 In the section "Node feature learning based on DMTT", the authors mentioned that "a new multi-scale neighbor topology embedding matrix  for each node is reconstructed through AMNE based on ." Please explain the reconstruction process of .  
  
  
2 Both Eq. (9) and Eq. (13) use the symbol , and Eq. (9) clarifies that  denotes the Hadamard product. I suggest that the meaning of  is given near Eq. (13) as well. It would be helpful for the authors to restate the definition of  in the description of Eq. (13) to improve readability.  
  
  
3 The authors proposed a strategy for adaptively fusing pairwise local features and global features. Please provide the final values of  and  to demonstrate which kind of features contributes more significantly.  
  
  
4 In the "Parameter settings" section, the number of layers L in DMTT is set to 2. The authors are encouraged to present results for different values of L to assess the impact of layer depth on model performance.  
  
  
5 The author can increase the citation of relevant articles to improve their readability, such as 10.1109/JBHI.2025.3561197, 10.1109/TBDATA.2023.3334673, 10.1021/acs.jcim.4c02250 and 10.1007/s11432-024-4098-3. The conclusion section of the paper should include a future work for this method.  
  
Reviewer: 10  
  
Comments to the Corresponding Author  
(There are no comments. Please check to see if comments were included as a file attachment with this e-mail or as an attachment in your Author Center.)