

## **Reviewer Comments**

### **Improved deep learning image classification algorithm based on Swin Transformer v2**

This paper presents an improved deep-learning image classification algorithm based on Swin Transformer for feature extraction and classification. The authors make several significant contributions to the proposed approach, which are worthy of review.

1. The authors proposed Swin Transformer, a novel block-structured architecture, to model the image input and effectively combine long-range dependencies for feature extraction. This novel approach to image recognition is an interesting development in the field.
2. The authors used Haarlike features and deep convolution neural networks (DCNN) to extract features, allowing for better feature representation and improved accuracy. This combination of techniques is interesting and innovative and could have interesting implications for further work in the image classification field.
3. The proposed method is better than many state-of-the-art methods in accuracy. The experiments conducted by the authors to test the algorithm were thorough and well-outlined.
4. The authors provide several performance metrics for their proposed algorithm, including accuracy, precision, recall, and F1-score. These metrics give an essential insight into the effectiveness of the proposed algorithm.
5. Using the Swin Transformer as a backbone network is both novel and promising. Proper evaluations should be made to measure the scalability of the proposed algorithm on larger datasets.
6. The proposed solution lacks empirical evidence of its testing on real-world datasets and the potential optimization techniques that can enhance the performance of the proposed model. The authors should present more evidence of applying the proposed algorithm to real-world datasets.
7. The authors have outlined the limitations in the Discussion section of their work; however, they have failed to provide a list of possible future works. A list of likely future works should be provided for improving the model's accuracy and exploring the potential for enhancement of image classification.
8. The authors should provide a more comprehensive explanation of the motivation for building the Swin Transformer architecture. Further, an approach with more elaborate theoretical insights should be undertaken for building and optimizing the model.

Overall, this is a well-written paper that addresses many of the critical issues in the field of image classification. The authors contributed significantly to the literature on image classification, and the paper should be accepted for publication.