Based on the reviewers’ suggestions, we highlight the results of the empirical study in the abstract as follows:

|  |
| --- |
| Lifelong learning for whole-slide images (WSIs) poses the challenge of training a unified model to perform multiple WSI-related tasks, such as cancer subtyping and tumor classification, in a distributed, continual fashion. This is a practical and applicable problem in clinics and hospitals, as WSIs are large, require storage, processing, and transfer time. Training new models whenever new tasks are defined is time-consuming. Recent work has applied regularization‑ and rehearsal‑based methods to this setting. However, the rise of vision‑language foundation models that align diagnostic text with pathology images raises the question: *are these models alone sufficient for lifelong WSI learning using zero‑shot classification, or is further investigation into continual‑learning strategies needed to improve performance?***The empirical study demonstrates that a well-pretrained pathology vision-language foundation model, when used with a simple zero-shot approach, can achieve competitive performance compared to training-based rehearsal and regularization-based continual learning methods**. To our knowledge, this is the first study to compare conventional continual‑learning approaches with vision‑language zero‑shot classification for WSIs. Our source code and experimental results will be available at https://github.com/caodoanh2001/ZeroSlide. |