Accuracy Comparison of Lightweight Vision-Language Models for Multimodal Fake News Detection

E.M.C.P. Ekanayake*, D. Herath

University of Peradeniya, Peradeniya, 20400, Sri Lanka
**panchaliekanayake1999@gmail.com*

Fake news shared on digital platforms increasingly combines textual and visual content to spread misleading information. The Indian Fake News Dataset (IFND) is a multimodal benchmark for fake news detection in the Indian news domain. The use of lightweight vision-language models for multimodal fake news detection on the IFND dataset has not yet been explored. We focus on lightweight vision-language models because fake news detection on social media platforms requires models that can process large amounts of multimodal content efficiently and with reasonable computing cost. In this work, we conduct a comparative analysis of lightweight vision-language models for binary fake news detection on IFND. Each sample is represented as a caption-image pair, and the model predicts whether the multimodal content is true or fake. The study is conducted on a subset of IFND containing 25,224 samples. Our experiments compare the performance of six lightweight vision-language models on this benchmark. Among the evaluated models, InstructBLIP achieved the best test accuracy of 72%, while Qwen-VL showed the lowest test accuracy of 33%. These results show that lightweight vision-language models can be applied for multimodal fake news detection, although their performance differs considerably across model architectures. Overall, this work presents a benchmark of lightweight vision-language models on IFND and highlights their potential for practical multimodal misinformation detection.

Keywords: fake news detection, multimodal misinformation, lightweight vision-language models, IFND