Multimodal Misinformation Detection in a South African Social Media Environment

A. De Jager[1], V. Marivate[1 [0000−0002−6731−6267]], and A. Modupe[1 [0000−0002−9732−6466]]

Department of Computer Science,
University of Pretoria, Pretoria 0028, South Africa
amicadejager@gmail.com, vukosi.marivate,abiodun.modupe@up.ac.za

Abstract. The world is witnessing a growing epidemic of misinformation. Misinformation can have severe impacts on society across multiple domains: including health, politics, security, the environment, the economy and education. With the constant spread of misinformation on social media networks, a need has arisen to continuously assess the veracity of digital content. This need has inspired numerous research efforts on the development of misinformation detection (MD) models. However, many models do not use all information available to them and existing research contains a lack of relevant datasets to train the models, specifically within the South African social media environment. The aim of this paper is to investigate the transferability of knowledge of a MD model between different contextual environments. This research contributes a multimodal MD model capable of functioning in the South African social media environment, as well as introduces a South African misinformation dataset. The model makes use of multiple sources of information for misinformation detection, namely: textual and visual elements. It uses bidirectional encoder representations from transformers (BERT) as the textual encoder and a residual network (ResNet) as the visual encoder. The model is trained and evaluated on the Fakeddit dataset and a South African misinformation dataset. Results show that using South African samples in the training of the model increases model performance, in a South African contextual environment, and that a multimodal model retains significantly more knowledge than both the textual and visual unimodal models. Our study suggests that the performance of a misinformation detection model is influenced by the cultural nuances of its operating environment and multimodal models assist in the transferability of knowledge between different contextual environments. Therefore, local data should be incorporated into the training process of a misinformation detection model in order to optimize model performance.