

Review of: "Comparative Analysis of Machine and Deep Learning Techniques for Text Classification with Emphasis on Data Preprocessing"

Mathias Fonkam¹

¹ ABTI American University of Nigeria

Potential competing interests: No potential competing interests to declare.

The title of this work and the broad goal of applying machine learning and deep learning techniques to the important task of unstructured text classification is very appealing given the potential use cases and application domains. The work is even more appealing given the emphasis in the title on data preprocessing, since this is an all-important part of any machine learning, and by extension, deep learning work. The work compares and evaluates a number of deep learning techniques on a familiar dataset - the Titanic dataset - concluding that the BiLSTM is the most precise model compared to other models. I will suggest considering the following questions that come to mind:

- Can you discuss why the Titanic dataset was chosen for this study and whether it is representative of medical text classification tasks?
- Can you provide examples of text classification tasks in the medical and other fields that highlight the limitations of using the Titanic dataset?
- Could you provide specific examples of how different preprocessing techniques might impact the performance of the BiLSTM model in this study?
- What are some specific preprocessing techniques that could be added to improve the performance of the deep learning models in this study?