MSR59

Advancing Systematic Literature Reviews: The Integration of AI-Powered NLP Models in Data Collection Processes

Pankaj Rai[1], Rajdeep Kaur[1], Shubhram Pandey[1], Sumeet Attri[1], Gagandeep Kaur[1], Barinder Singh[2]
[1]Pharmacoevidence, Mohali, India, [2]Pharmacoevidence, London, UK

Pharmaco Evidence

CONCLUSION

The NLP-driven strategy employed in automating SLR screening proved effective, with the BERT algorithm exhibiting the highest accuracy among all models studied. This automation reduces manual workload, enhancing efficiency and representing a significant improvement in optimizing the SLR process.

INTRODUCTION

- A systematic literature review (SLR) involves extensively evaluating the quality of previous research and presenting key findings on a topic. However, the process is time-consuming and sometimes results in outdated findings[1]
- Incorporating natural language processing (NLP) models[2-5] to automate SLRs signifies groundbreaking progress in research methodology
- These models harness the power of artificial intelligence (AI) to navigate through extensive content efficiently, streamlining the SLR process undertaken by human reviewers

OBJECTIVES

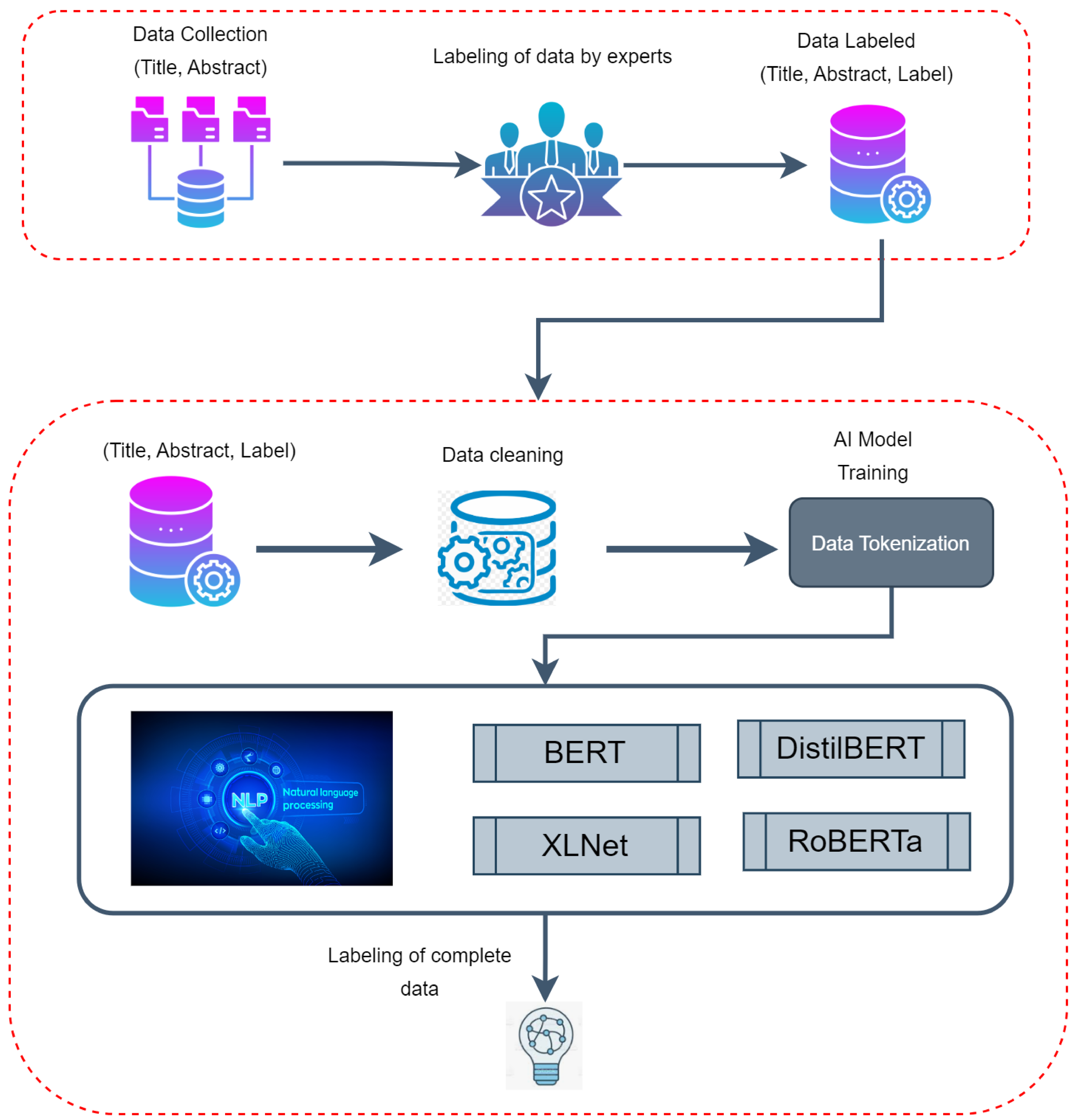
- This SLR aimed to assess the efficiency of different NLP models, including BERT, DistilBERT, RoBERTa, and XLNet, based on semantic analysis of titles and abstracts

METHODS

- Python-based NLP models (BERT, DistilBERT, RoBERTa, and XLNet) were developed to enhance the efficiency of screening literature for SLRs
- A domain expert with over a decade of experience manually screened the title and abstracts of the data sample for model training and improvement
- Subsequently, to effectively capture contextual relationships within the data, texts underwent tokenization using tokenizers specific to the models
- The models' performance was validated using the remaining data, which constituted previously unseen data
- To address the class imbalances, a random oversampling to ensure a balanced dataset was employed
- To ensure a thorough and accurate assessment of the models' capabilities, the training set was subdivided through K-Fold Cross-Validation (K=5), enhancing robustness in the evaluation; the model results were compared using the performance metrics
- **Figure 1** provides a comprehensive depiction of the entire process
- **Metrics:** Accuracy and Sensitivity are calculated using confusion matrix values (**Figure 2**), i.e., True Positive, False Positive, False Negative, and True Negative using the following formulas:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \qquad \text{Sensitivity} = \frac{TP}{TP+FN}$$

Figure 1: First stage screening using NLP models



RESULTS

- Across the various NLP models considered for title and abstract-based screening, BERT showcased better performance in the validation phase, with an accuracy of 90.05% and a sensitivity of 84.16%, surpassing other models (**Figure 3**)
- DistilBERT closely followed with competitive results, achieving an accuracy and sensitivity of 88.90% and 75.25%, respectively (**Figure 3**)
- XLNet performed well, securing an accuracy of 87.24% and a sensitivity of 81.19% (**Figure 3**)
- Nevertheless, RoBERTa demonstrated a marginally lower accuracy of 78.34%, coupled with a sensitivity of 82.10%, suggesting a relative performance dip (Figure 3)

Figure 2: Confusion matrix of different models

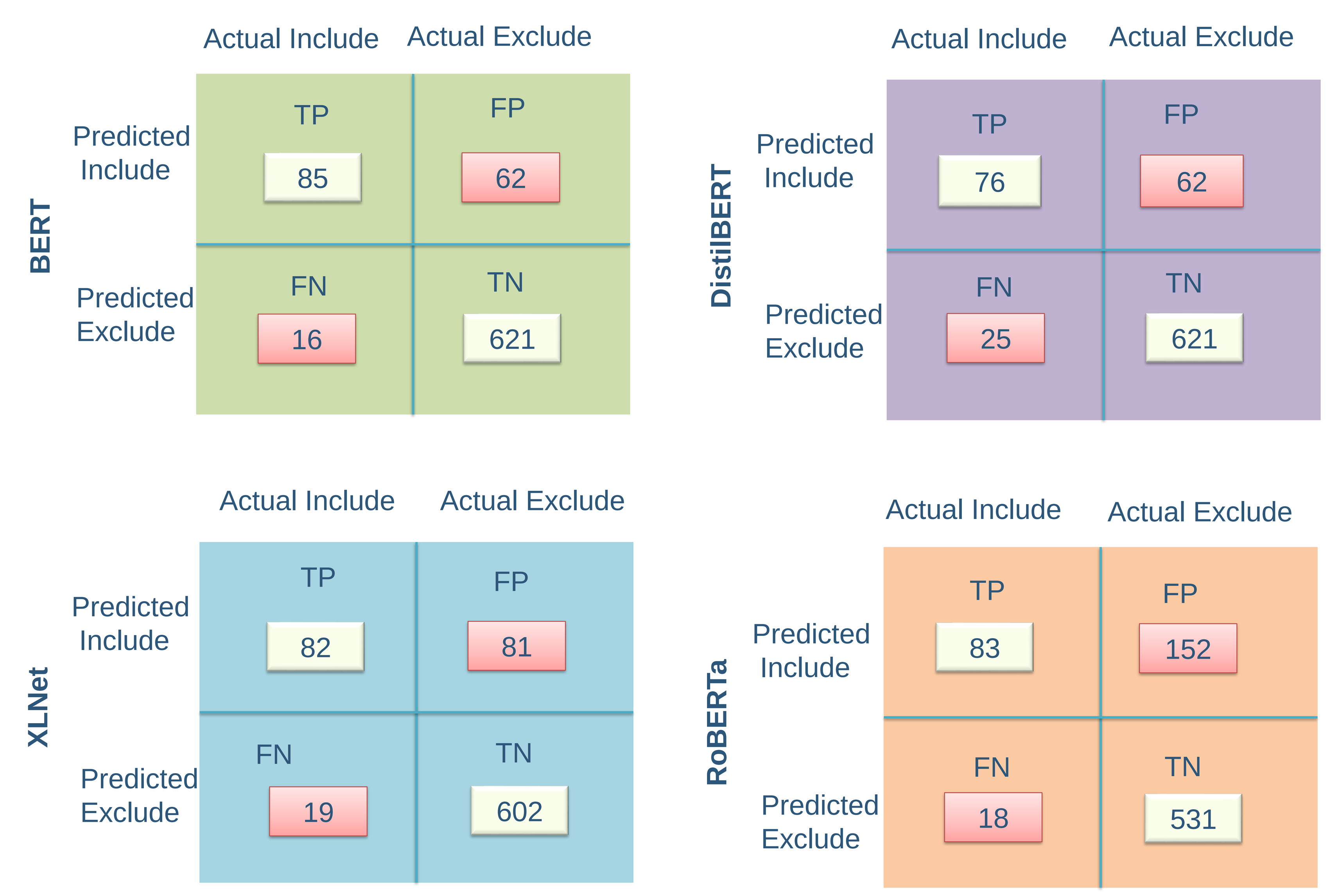
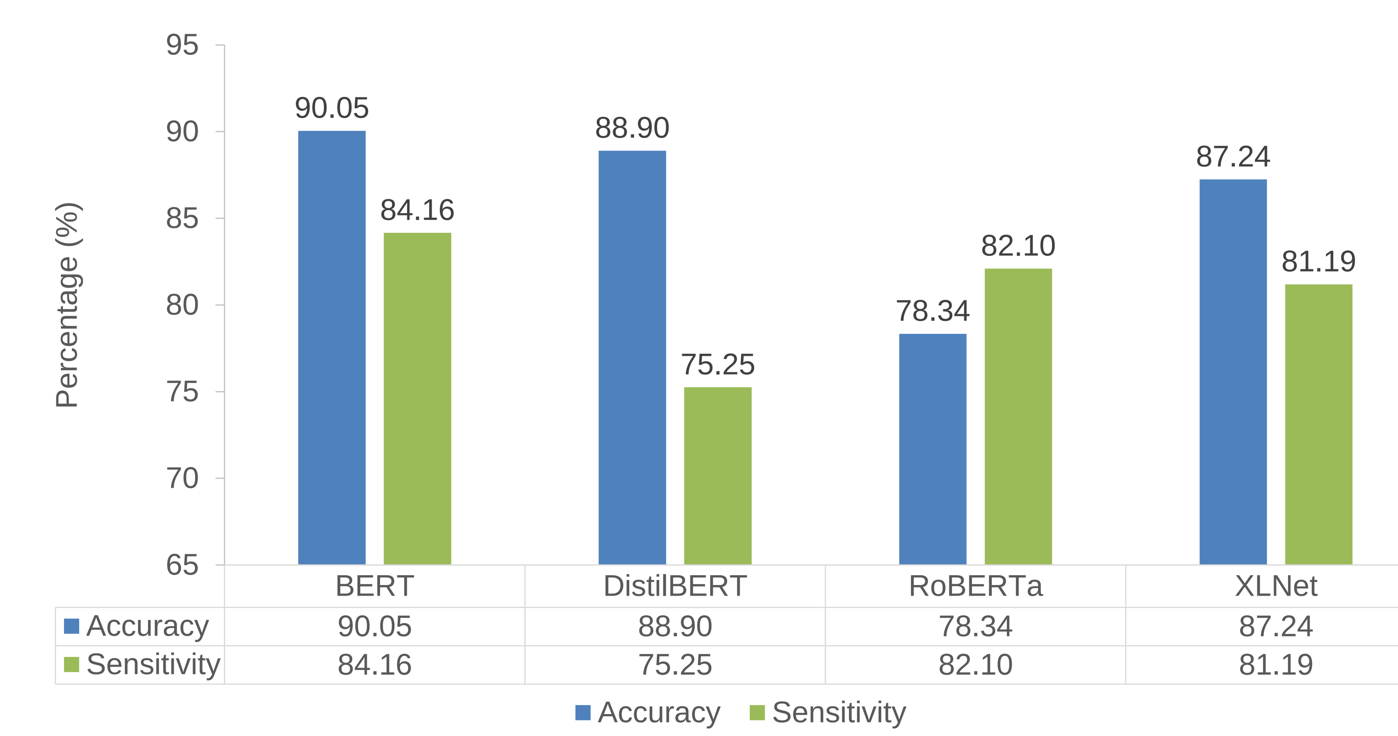


Figure 3: Comparison of performance of different NLP models



REFERENCES

1. Aum, S., & Choe, S. (2021). srBERT: automatic article classification model for systematic review using BERT: Systematic reviews, 10 : 1-8.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
3. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108. Aum, S., & Choe, S. (2021).
4. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
5. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems

Disclosure

PR, RK, SP, SA, GK, and BS, the authors, declare that they have no conflict of interest

For further queries, please contact: Barinder.Singh@Pharmacoevidence.com

ISPOR Improving healthcare decisions