Over the past few decades, the rate of scientific publication has increased dramatically. At the same time, the number of retractions and failed replications of studies have also increased. This replication crisis across many scientific disciplines has heightened scrutiny of scientific papers and their reported results. However, the incentive for researchers to pursue replication of prior research is limited and are typically outweighed by the costs. As such, there is a need for computational methods for prioritizing replication efforts on scientific findings that are most dubious (i.e., a high estimated probability of retraction) and highest impact (i.e., a large number of citations). The goal of this final project is to address the first desiderata by building a classifier for detecting dubious research articles from their titles and abstracts.

**Background**

- Published vs Retracted/Replicated
- Number of articles over time

**Methods**

- SciBERT transformer model
  - 1.14 million full-text research papers
  - 20% CS + 80% Biomedical
  - Specially tailored SciVocab
- Random forest baseline
  - Trained on TF-IDF vectors

**Experiments**

- **Fine-tuning:** Training updates to all SciBERT weights
- **Frozen weights:** Training updates only to classifier head

**Parameters:**
- 20 epochs with early stopping; optimizer = SGD; loss = BCE, lr=2e-5; decay = 0.001; batch=32

**Conclusions**

- Machine learning models can classify retracted papers from matched negative samples using title and abstract text.
- Including additional features may increase performance but could also introduce bias.
- In the future, it is necessary to evaluate the efficacy of the predicted probability of retraction (in conjunction with scientific impact metrics) for prioritizing replication efforts.

---


---