

# Quick, get me a Dr.BERT: Automatic Grading of Evidence using Transfer Learning

Our approach got us third place in the ALTA 2021 Shared Task Challenge.

SPOILER ALERT: Grading medical evidence can be tricky!

Pradeesh Parameswaran  
Andrew Trotman  
Veronica Liesaputra  
David Eyers



## 01 INTRODUCTION

- COVID-19 pandemic raises importance of Evidence-Based Medicine (EBM)
- Strength Of Recommendation Taxonomy (SORT) is a popular EBM framework

## 02 RESEARCH QUESTIONS

- Can we use BERT to accurately perform SORT?
- Can we use author's and journal features to improve the scores?

## 03 METHODOLOGY

Two-phase approach

- Phase 1—BioBERT (*abstract*)
- Phase 2—SVM Classifier (*author/journal features*)

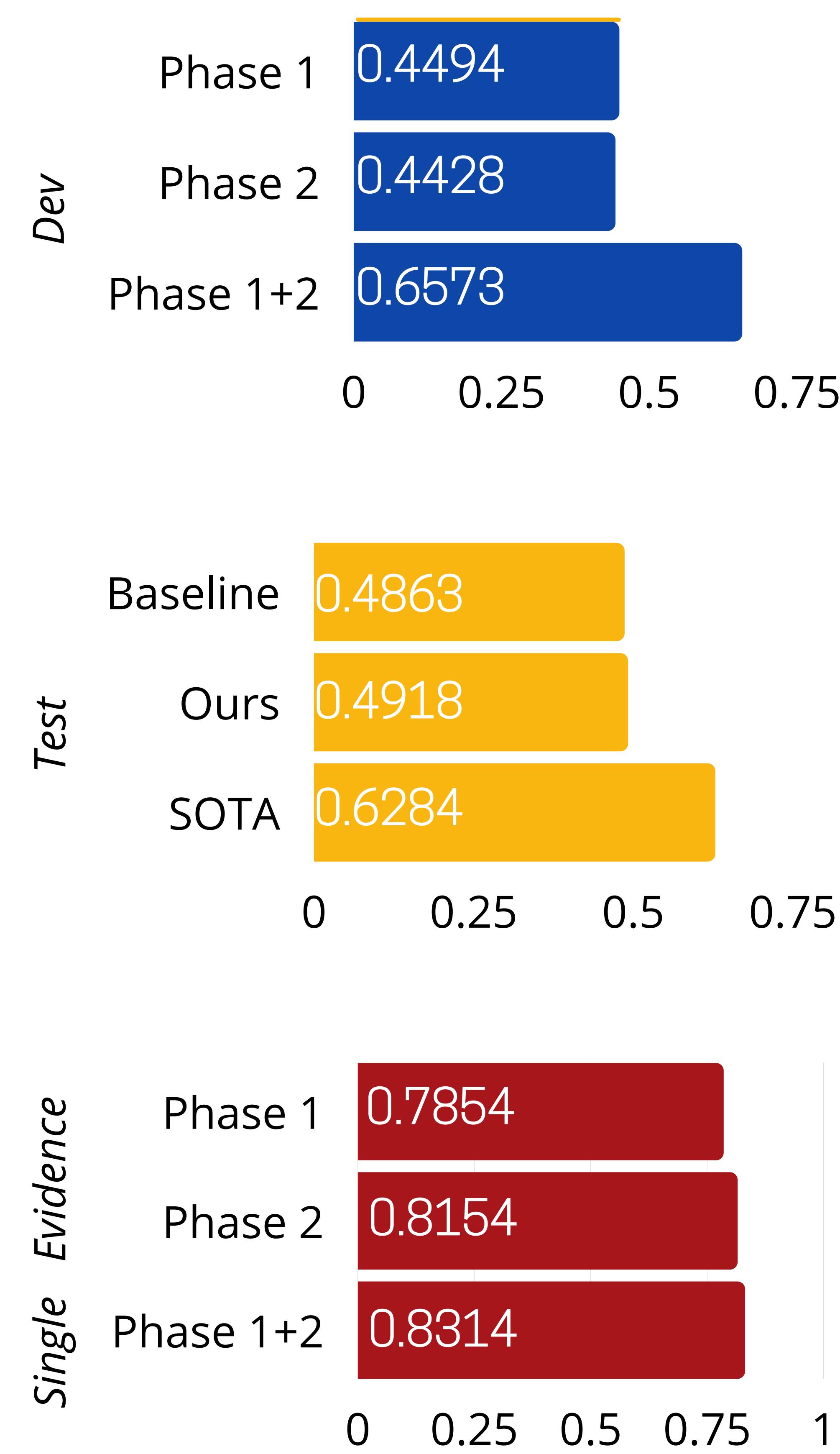
### 3 Grades in SORT

- **A** (Consistent Findings)
- **B** (Inconsistent Findings)
- **C** (Expert's Opinion)



## 04 RESULTS

Evaluated on CodaLabs using Accuracy Score



## 05 FAILURE ANALYSIS

### What worked well:

- Classifying a single piece of evidence
- Average h-index of all authors > h-index of first author
- Substituting medical terms with generic token

### What didn't work:

- Deriving Final Grade
  - Can't average grades
    - Two studies may have '**A**' grade but ≠ conclusions = '**B**'
    - RCT > Opinion
- Distinguishing '**A**' and '**B**' is HARD!
  - No consensus on RCT Size
  - The style of writing is identical

## 06 CONCLUSION

BERT + SVM Classifier alone is not enough to beat past approaches for automating SORT.

### Future work:

- Improve classification on multiple evidence
- Use document similarity techniques

