



Biomarker or Not: Automated Biomarker Detection in Alzheimer's Clinical Trial Text

Gaurab Subedi¹, Amanda Leisgang Osse^{2,3}, Andrew Hooyman⁴, Yadi Zhou⁵, Jefferson Kinney^{2,3}, Feixiong A. Cheng⁵, Jeffrey Cummings, MD, ScD^{2,3}, Jorge Fonseca, Ph.D.¹.

¹Department of Computer Science, University of Nevada, Las Vegas, Las Vegas, USA, ²Kirk Kerkorian School of Medicine, University of Nevada, Las Vegas, Las Vegas, NV, USA, ³Chambers-Grundy Center for Transformative Neuroscience, Las Vegas, NV, USA, ⁴Physical Therapy, Crean College of Health and Behavioral Sciences, Chapman University, Orange, CA, USA, ⁵Cleveland Clinic, Cleveland, OH, USA

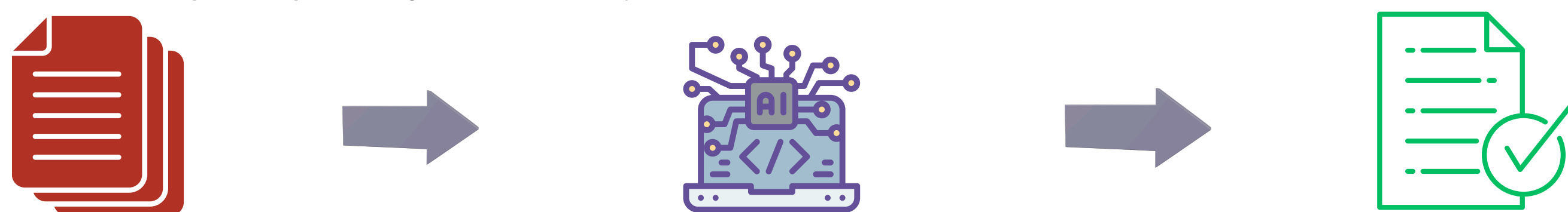
Funded by the Alzheimer's Drug Discovery Foundation (ADDF).



1 Introduction

Why this matters

- This work is part of the larger ADBMO effort to accelerate biomarker annotation in Alzheimer's disease clinical trials.
- Clinical trial records contain biomarker-related eligibility criteria and outcome measures, but manual review is slow and expert-intensive.
- We automatically identify biomarker presence and absence in trial text, narrowing expert review to only the passages that require annotation.



Goal:

- Benchmark AI models for biomarker mention detection to support faster trial triage and downstream expert annotation.

2 Methods

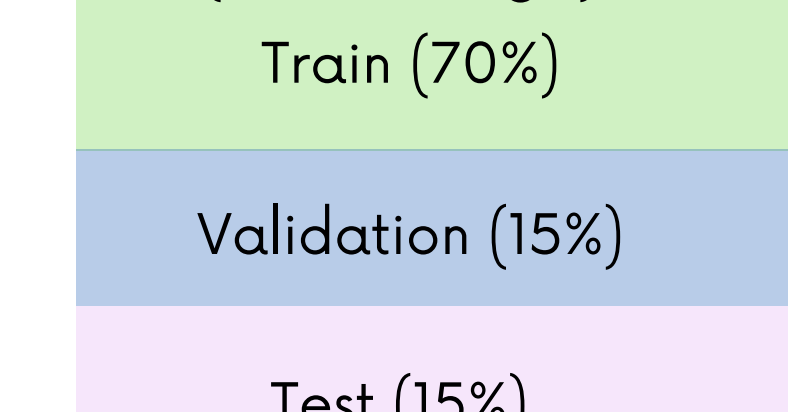
Clinical Trials.gov
AD Trials



1,900 Expert-Annotated Lines
from 234 Trials



Stratified Splits by NCTID
(No Leakage)



Feature Conditions

TF-IDF

TF-IDF + BioWordVec-200

TF-IDF + FastText-300

Classical Models

- Logistic regression
- SVM
- Random Forest
- XGBoost
- Decision Tree
- Artificial Neural Networks

Transformer Models

Fine-tuned (AdamW, class- weighted loss)

BERT-base, BioBERT, PubMedBERT, Bio+Clinical Bert, BioLinkBERT

Zero-shot (No Fine-tuning)

BART, FLAN-T5, BioGPT

Five-Fold Cross-Validation (Stratified by NCTID)

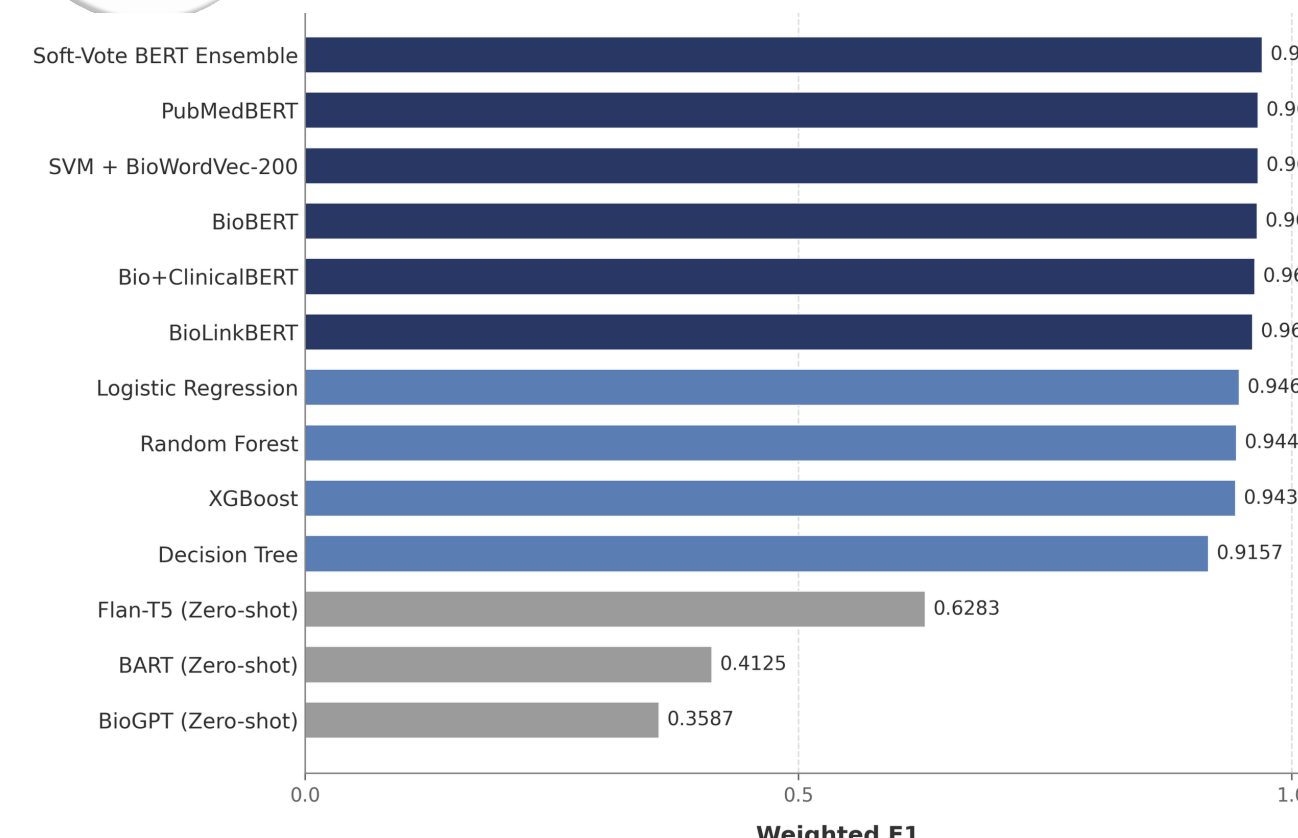
4 Discussion

Strong Classical Alternative: SVM with BioWordVec-200 nearly matched the best transformer while offering much faster inference, making it practical for large-scale trial triage.

Efficiency Matters: Fast inference enables scalable screening of trial records before expert annotation.

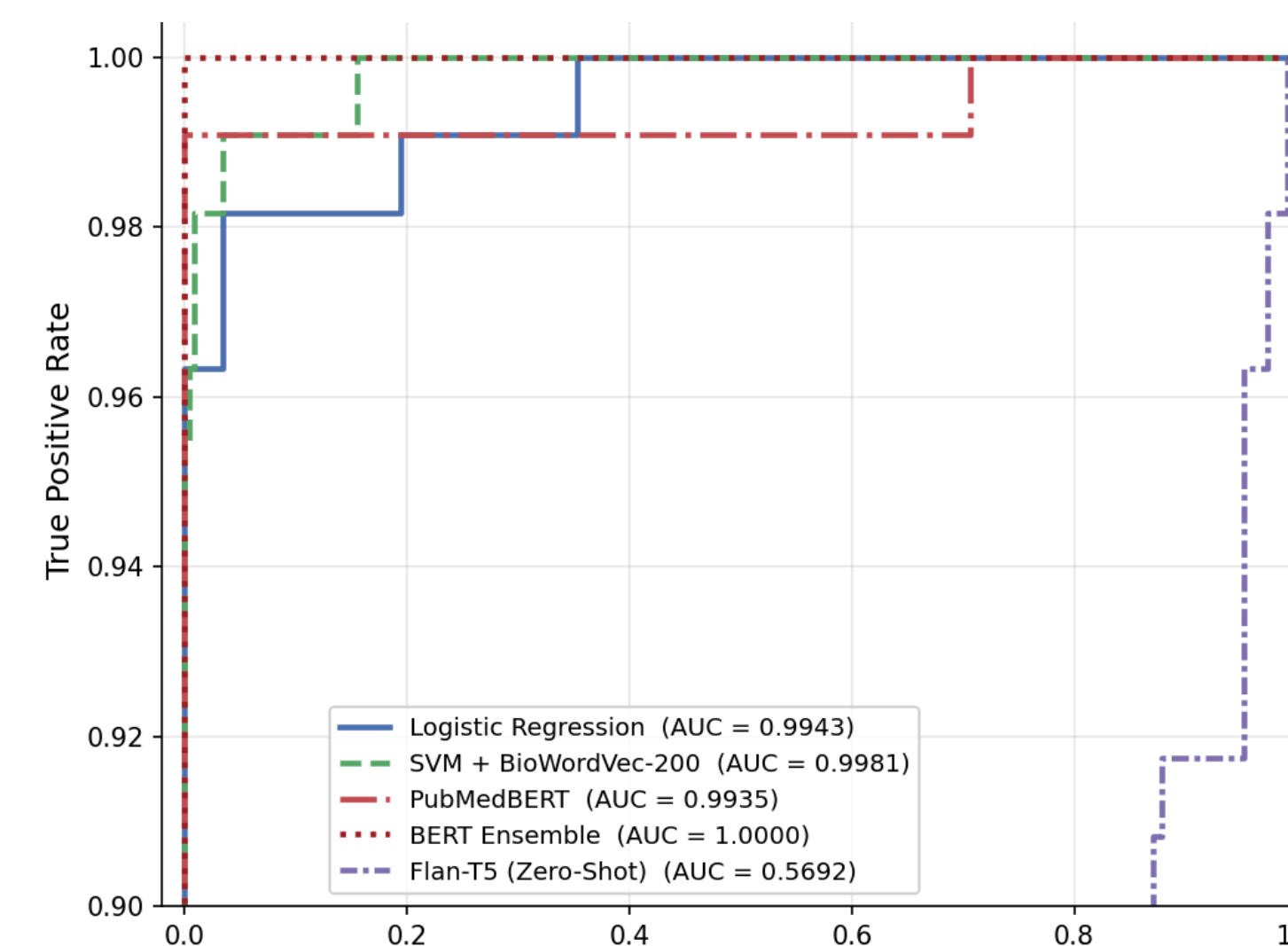
High Accuracy: Transformer models, especially PubMedBERT achieved best performance on both CV and test set.

Zero-Shot Limitations: General Foundational Model without task specific training performed poorly

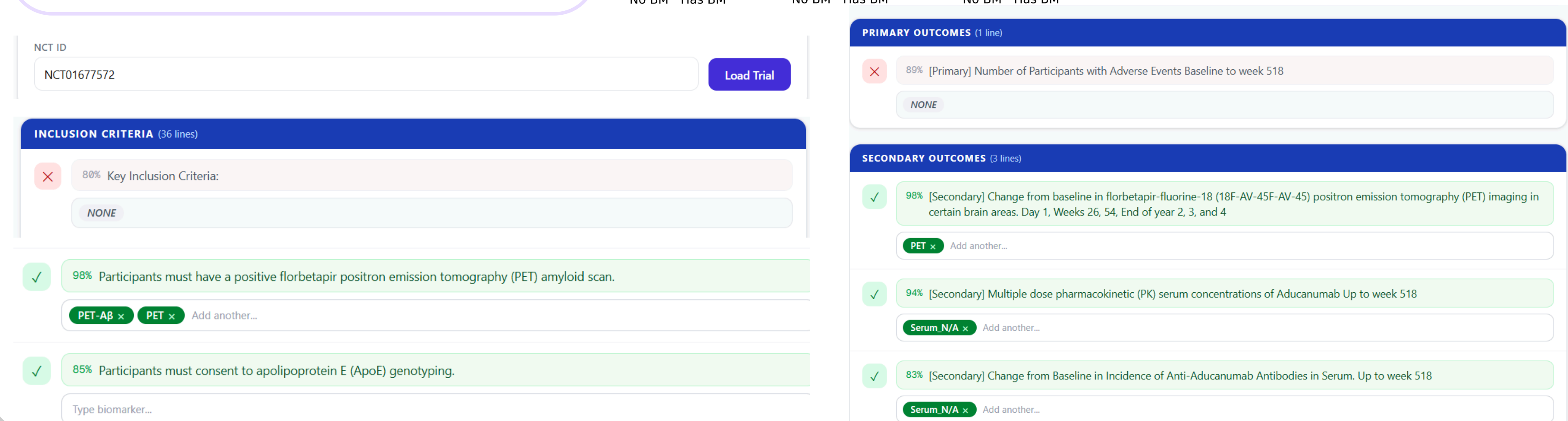
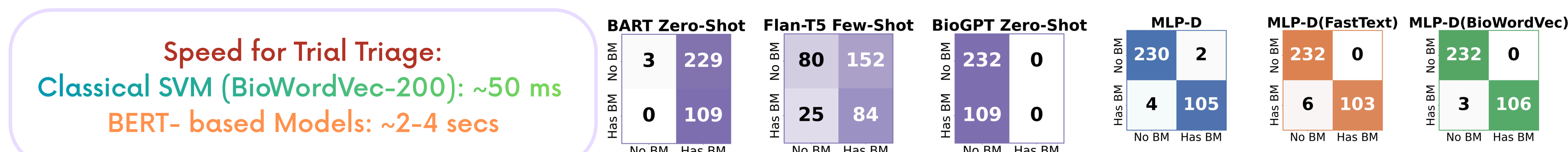


3 Results

- PubMedBERT was the strongest single model (CV F1 = 0.9660, AUC = 0.9947).
- The soft-vote BERT ensemble achieved CV F1 = 0.9702 and perfect AUC = 1.000 on the held-out test set.
- Logistic Regression reached CV F1 = 0.9469, while SVM with BioWordVec-200 embeddings (pre-trained on PubMed & MIMIC-III) reached 0.9658.
- Zero-shot models performed poorly, Flan-T5 was the best (F1=0.6283).
- On the test set, the four leading BERT models and Classical Ensemble converged at weighted F1 = 0.985 with biomarker-class F1 = 0.977.



Speed for Trial Triage:
Classical SVM (BioWordVec-200): ~50 ms
BERT-based Models: ~2-4 secs



5 Conclusion

- An SVM trained using BioWordVec embeddings produced comparable results to the transformer models but with faster inference. This shows that carefully training classical models with quality data can still be a competitive and efficient alternative to Transformer models.
- Task-specific training improved identification of biomarker-relevant text within the ADBMO pipeline.
- Automated biomarker flagging reduced annotation workload by directing experts to relevant text and filtering out filler.