CONTRACT NLI

ANLP Monsoon ‘24

Team 45, Fox 🦊

TEAM MEMBERS

Jhalak Banzal
2021101079

Ajit Srikanth
2021112023

Pranav Agrawal
2021101052

TA: Sidhi Panda

CONTRACT NLI

- NDA documenst from internet searches and EDGAR
- 17 Hypotheses for which annotation was conducted by a computational linguistic researcher
- Annotated spans (spans using Stanza)

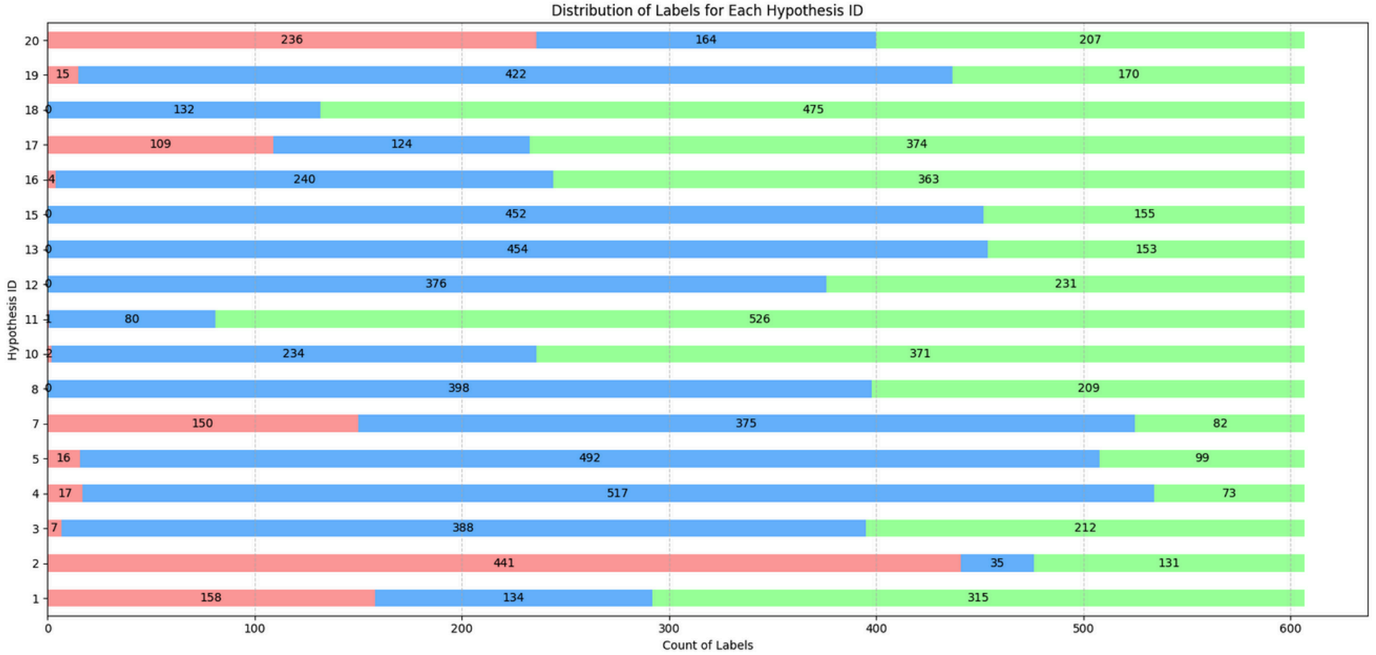
1. Evidence Identification
2. Natural Language Inference (NLI)

...
Confidential Information: means all confidential information (however recorded, preserved or disclosed) disclosed by a Party or its Representatives to the other Party and that Party's Representatives including but not limited to:
(a) the fact that discussions and negotiations are taking place concerning the Purpose and the status of those discussions and negotiations;
(b) the existence and terms of this Agreement;
(c) any information relating to:
(i) the business, affairs, customers, clients, suppliers, plans, intentions, or market opportunities of the Disclosing Party or of the Disclosing Party's Affiliates; and
(ii) the operations, processes, product information, know-how, designs, specifications, trade secrets, computer programs or software of the Disclosing Party or of the Disclosing Party's Affiliates; and
(d) any information or analysis derived from Confidential Information.
...

// denotes a span border

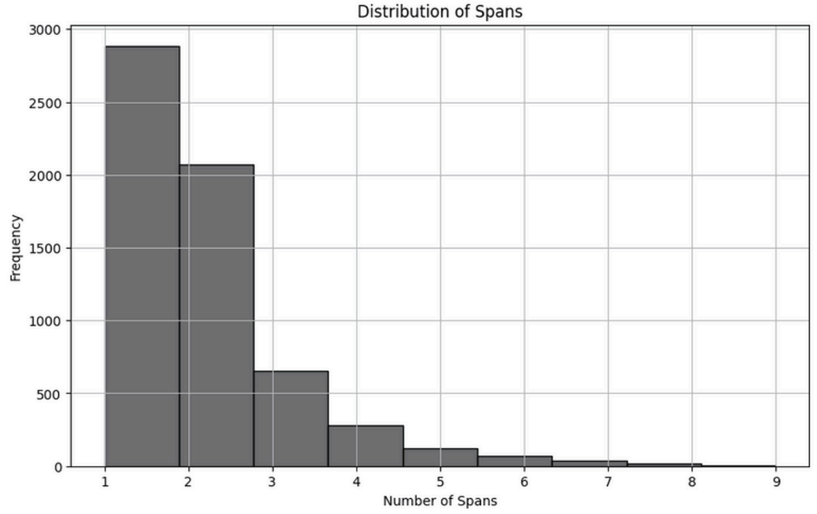
Examples of hypotheses:

Receiving Party shall not disclose the fact that Agreement was agreed or negotiated. (Evidence denoted with green highlight on upper half of text)	☑ Entailment ☐ Contradiction ☐ Not mentioned
Confidential Information shall only include technical information. (Evidence denoted with blue highlight on bottom half of text)	☐ Entailment ☑ Contradiction ☐ Not mentioned
Receiving Party shall not use any Confidential Information for any purposes other than the purpose(s) stated in Agreement. (Evidence does not exist when the hypothesis is not mentioned)	☐ Entailment ☐ Contradiction ☑ Not mentioned

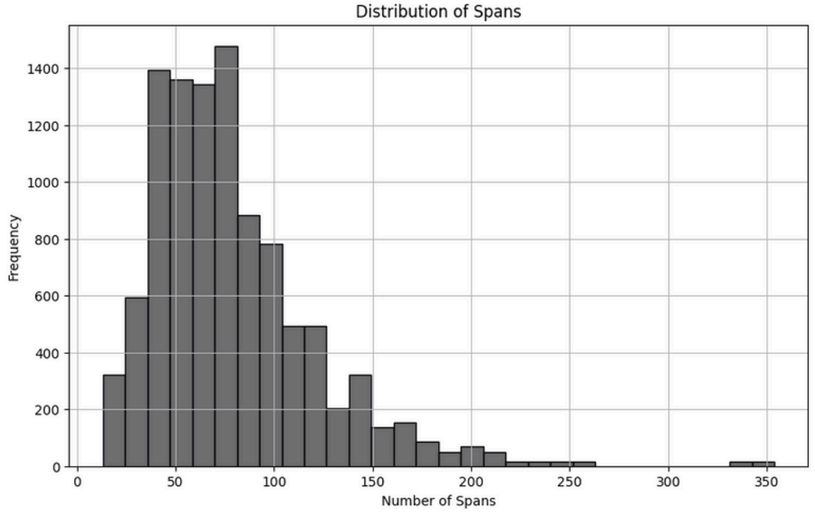


EDA

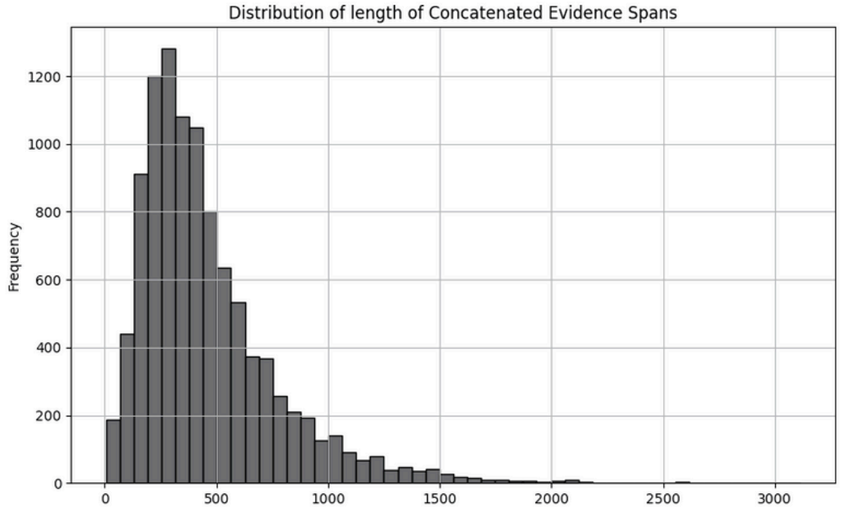
A document on average has 77.8 spans to choose evidence spans from. An average number of tokens per a document is 2,254, which is larger than maximum allowed context length of BERT (512).



Evidence Spans



All Spans



Concatenated Evidence Spans

RELATED WORK

DocInfer

Introduces CaseHoldNLI leveraging GANs and node pruning for efficient legal case entailment. Targets structured logical reasoning within domain-specific datasets.

Fast and Accurate Factual Inconsistency Detection Over Long Documents

Proposes the SCALE: Source Chunking Approach for scalable inconsistency evaluation in extensive texts.

Long Document Summarization

Head-wise positional strides allow different attention heads to focus on distinct segments.

Longformer

Combines local windowed, and global attention for specific tokens (e.g., CLS tokens).

BigBird

Implements hybrid sparse attention; Random, window-based, and global attention patterns.

Recurrent Memory Transformer

SPAN NLI BERT

Span-based tasks in document-level NLI face issues like span splitting across contexts, insufficient surrounding context, and the difficulty of detecting span boundaries and relevance simultaneously.

1

Dynamic Context Segmentation

Documents are split into overlapping contexts, ensuring spans are fully included and surrounded by sufficient context to loss of meaning.

2

Simplified Span Detection

[SPAN] tokens replace start-end token prediction, framing span relevance as a straightforward binary classification task.

3

Aggregation of Predictions

Span probabilities are averaged across contexts, and NLI predictions are weighted by span relevance to produce document-level results.

SPAN NLI BERT

1

Discontinuous Spans

In ContractNLI, 20% of spans are discontinuous, often spread across pages or sections, making them hard to capture in a single context.

Without overlapping segmentation, the model risks missing connections between span parts, reducing accuracy in evidence identification and NLI.

2

λ Tuning

The loss function combines evidence (Lspan) and NLI loss (LNLI) with a weight λ:

$$L = L\text{span} + \lambda L\text{NLI}$$

Instability arises as different tasks or documents need different λ values, complicating training.

PROPOSED APPROACH

TASK 1 Evidence Inference (EI)

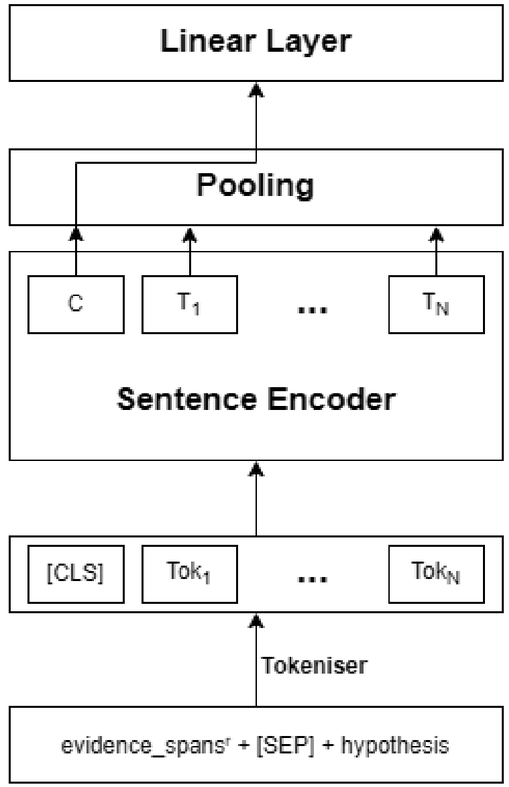
Identify which spans are evidence.

Inputs: [SEP]<hypothesis>

TASK 2 Natural Language Inference (NLI)

Use 2 models (for Entailment & Contradiction), to account for class imbalance in the dataset, to predict the label.

Input: <concatenated_pred_evidence_spans>[SEP]<hypothesis>



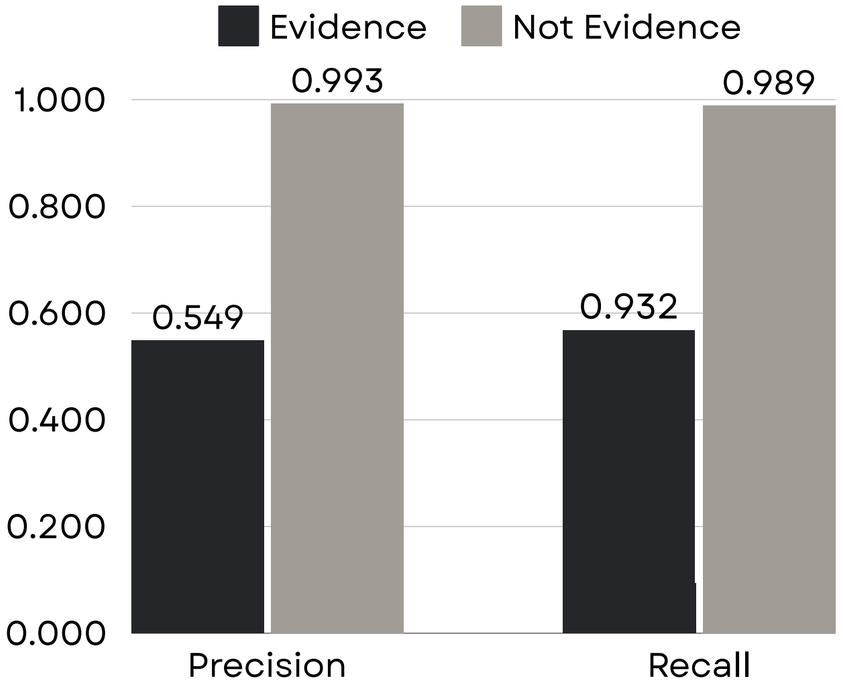
EXPERIMENTS

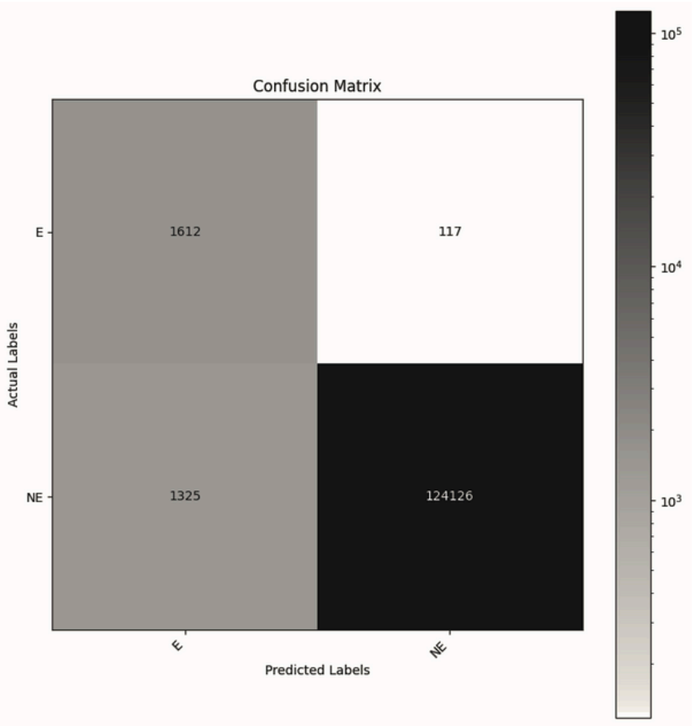
BERT	Masked Language Model (MLM), Next Sentence Prediction (NSP)
DistilBERT	Knowledge Distillation, MLM Only
RoBERTa	MLM with dynamic masking
ALBERT-base-v2	MLM, Sentence Order Prediction (SOP)
DeBERTa	MLM with disentangled attention, Replaced Token Detection (RTD)
BigBird	Block sparse attention instead of normal attention, 4096 seq len
ST:MiniLM-L6	Contrastive Learning Objective, distilled from RoBERTa
ST:MPNet	MLM with permutation-based pre-training for enhanced bidirectional context

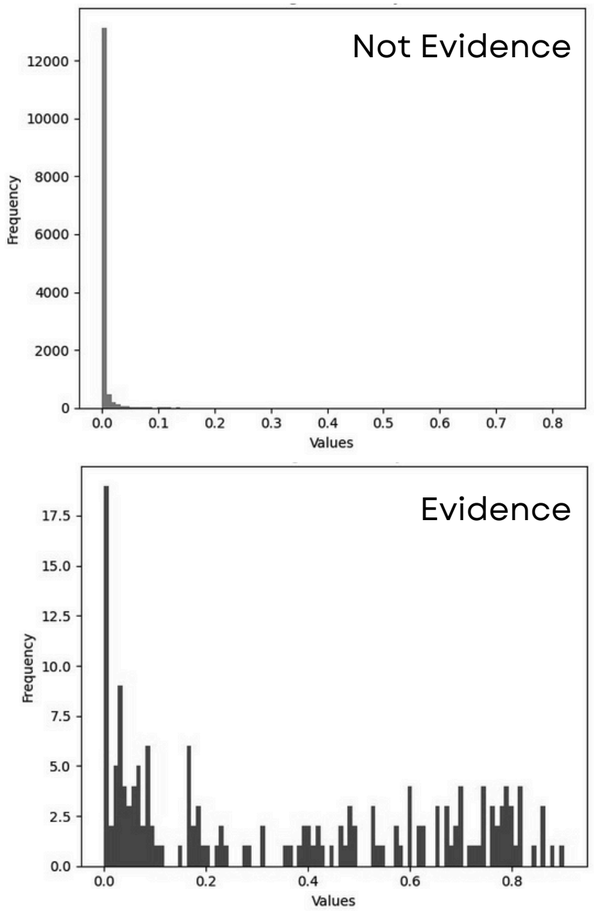
EI SUBTASK

Acc. : 98.87%

mAP : 0.771





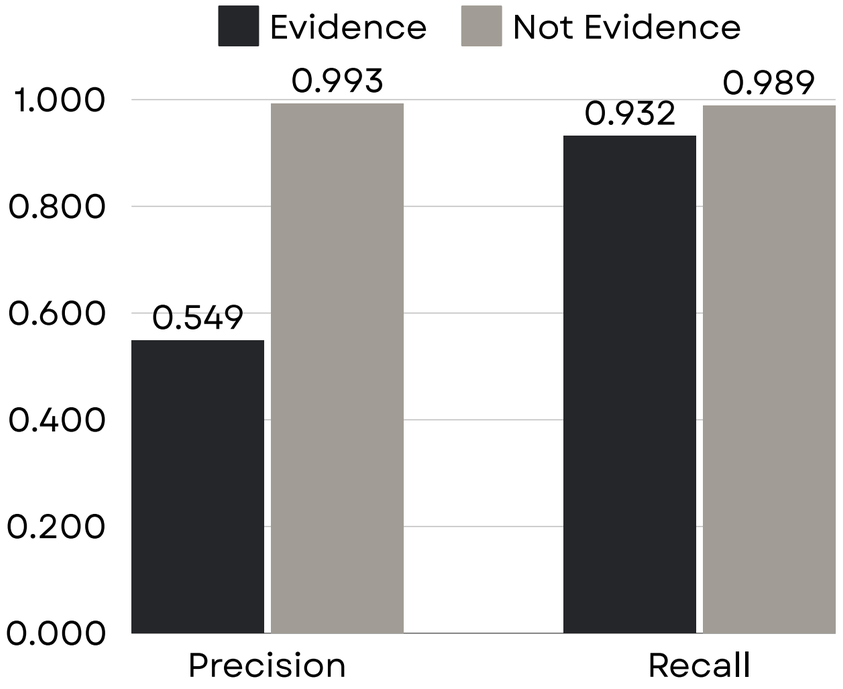


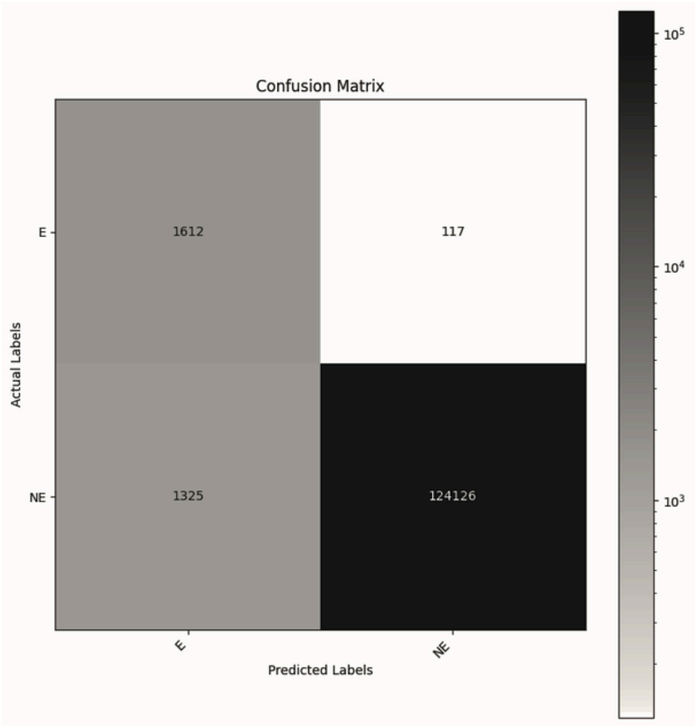
Confidence values to select threshold

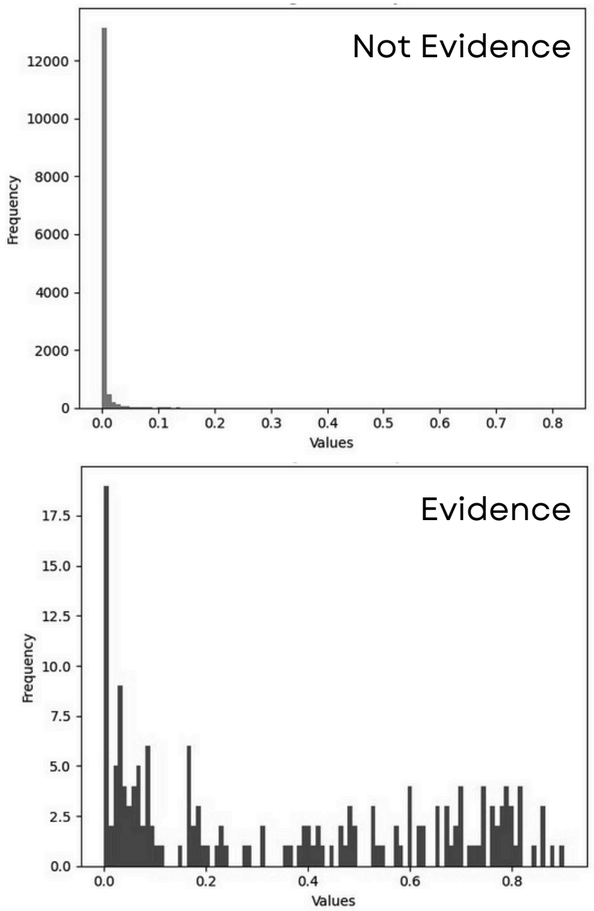
EI SUBTASK

Acc. : 98.87%

mAP : 0.771



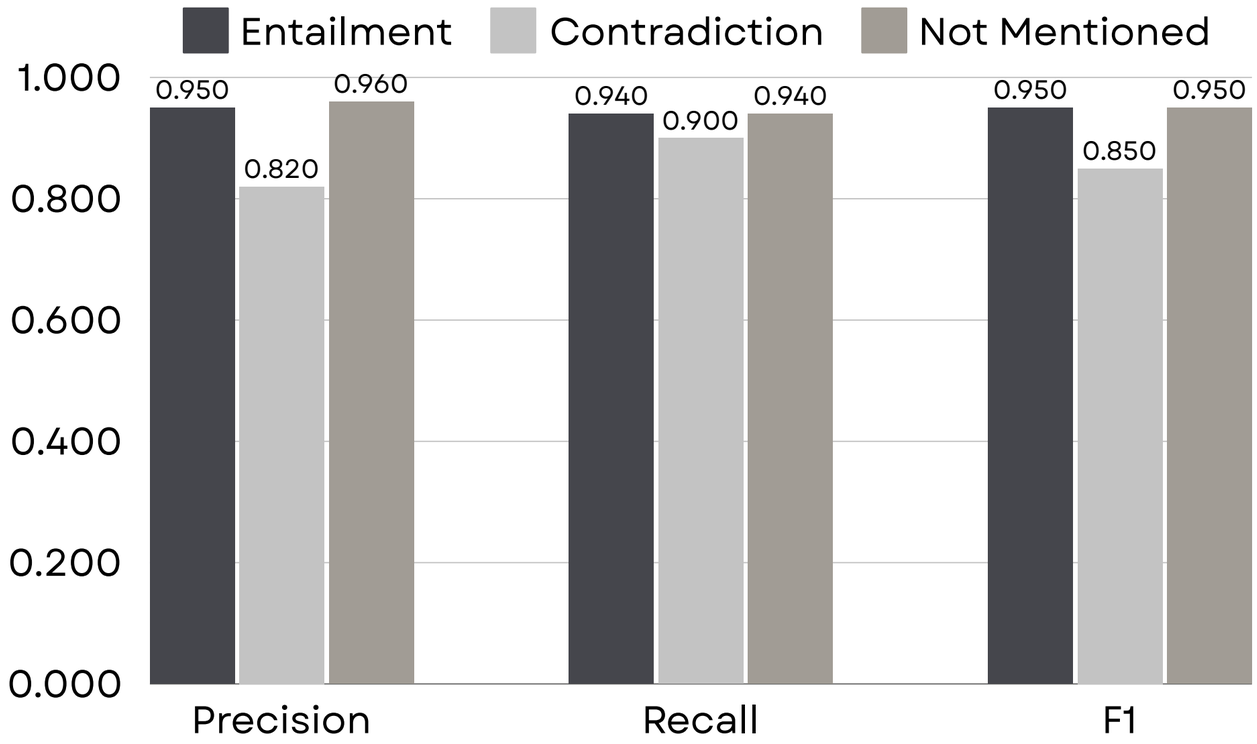


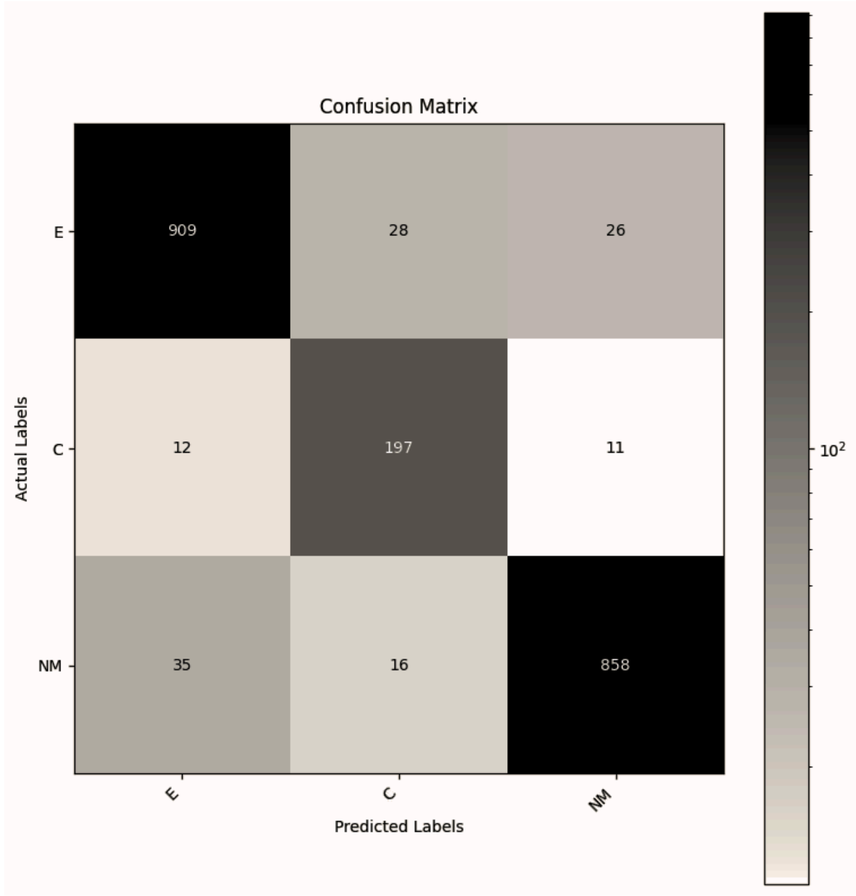


Confidence values to select threshold

NLI SUBTASK

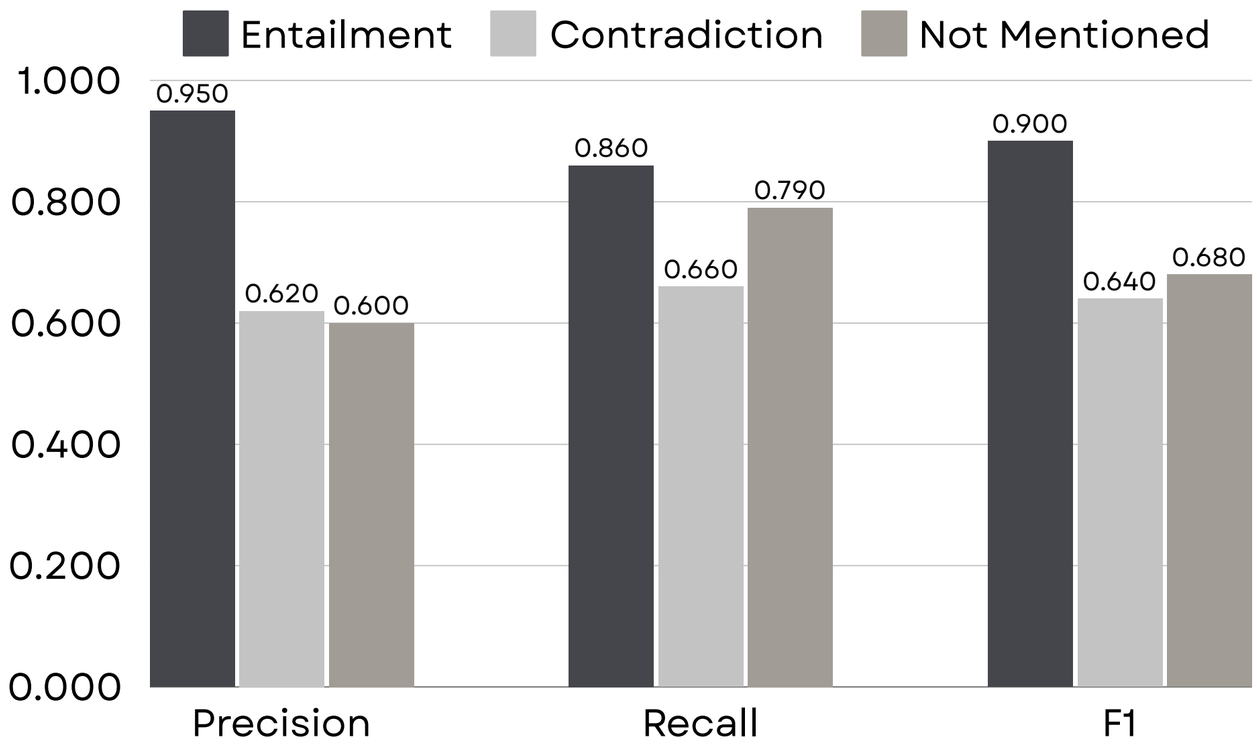
Acc: 93.88%

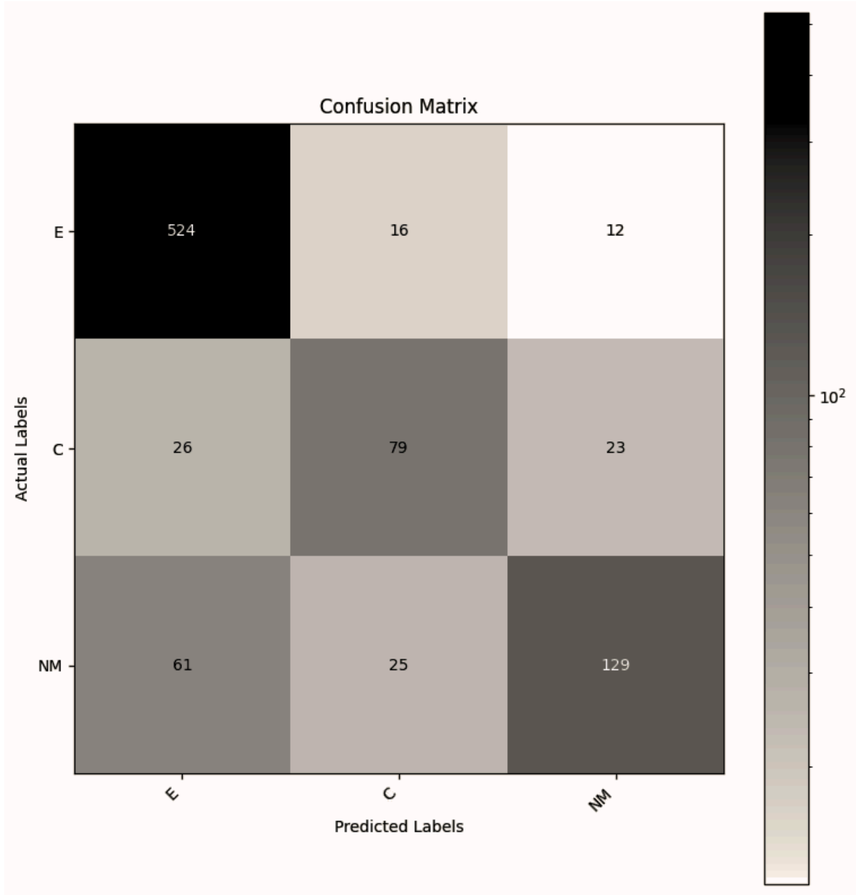




COMBINED NLI

Acc: 81.79%



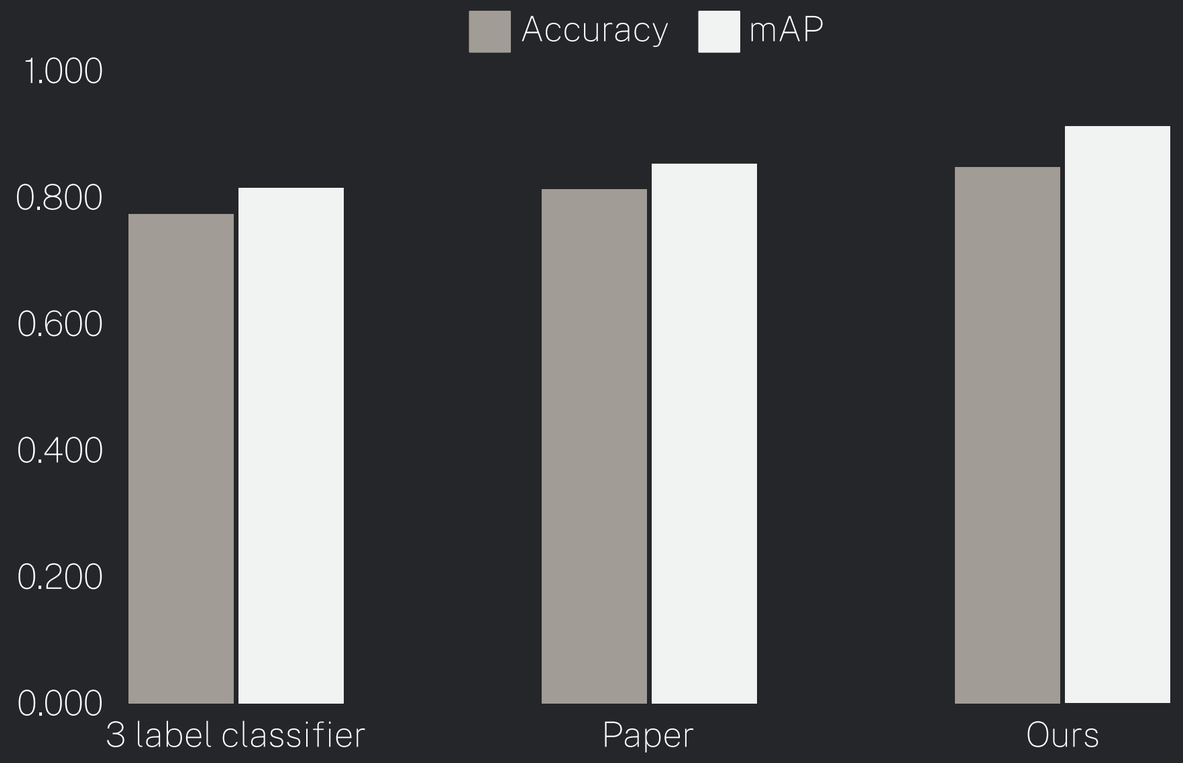


RESULTS

We see that our results were much better than the 3 label classifier, i.e. BERT-large.

Moreover, they also beat Span NLI BERT based on our experiments.

**methods like DocInfer produce superior results but are not in the scope of our project (clarification by Sidhi)*



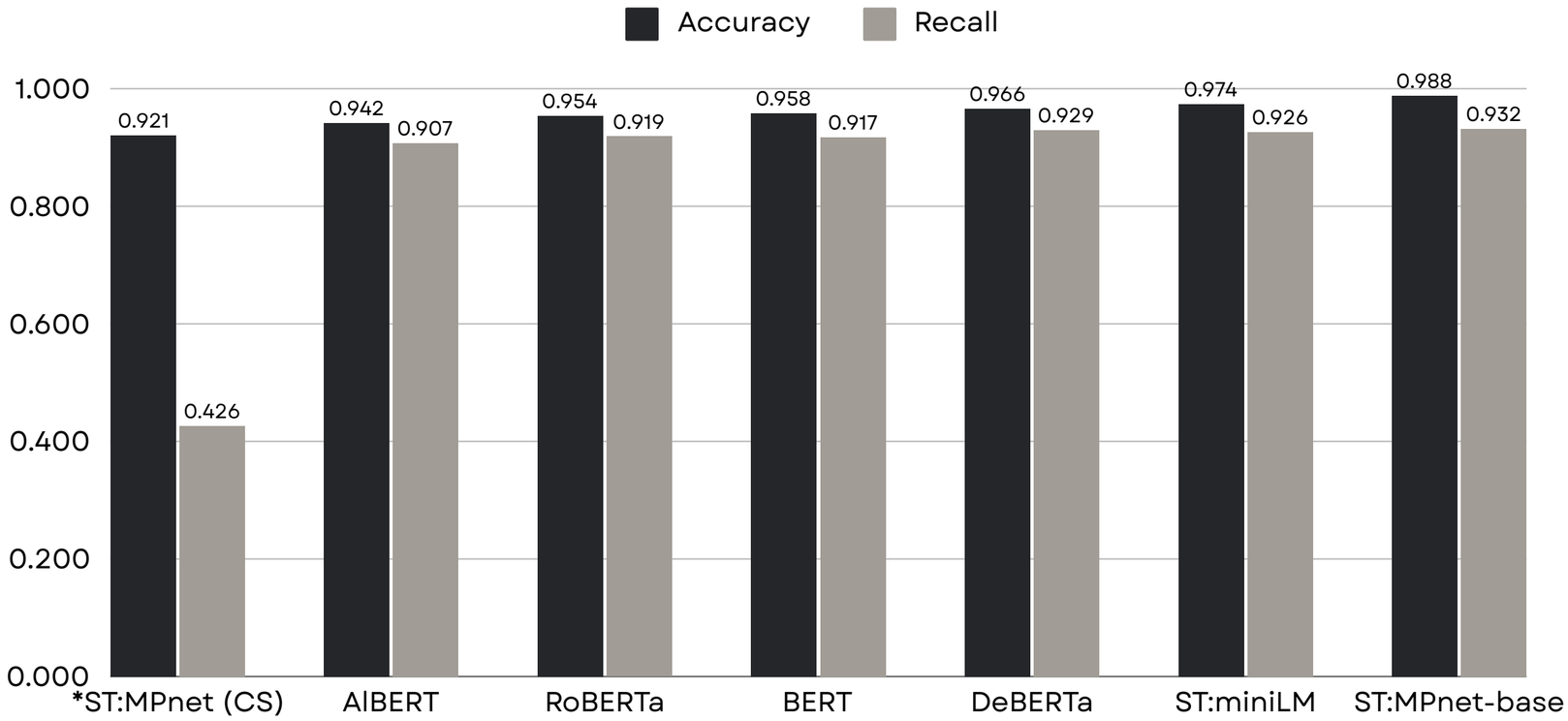
THANK YOU

(& special thanks to our TA, Sidhi Panda)

APPENDIX

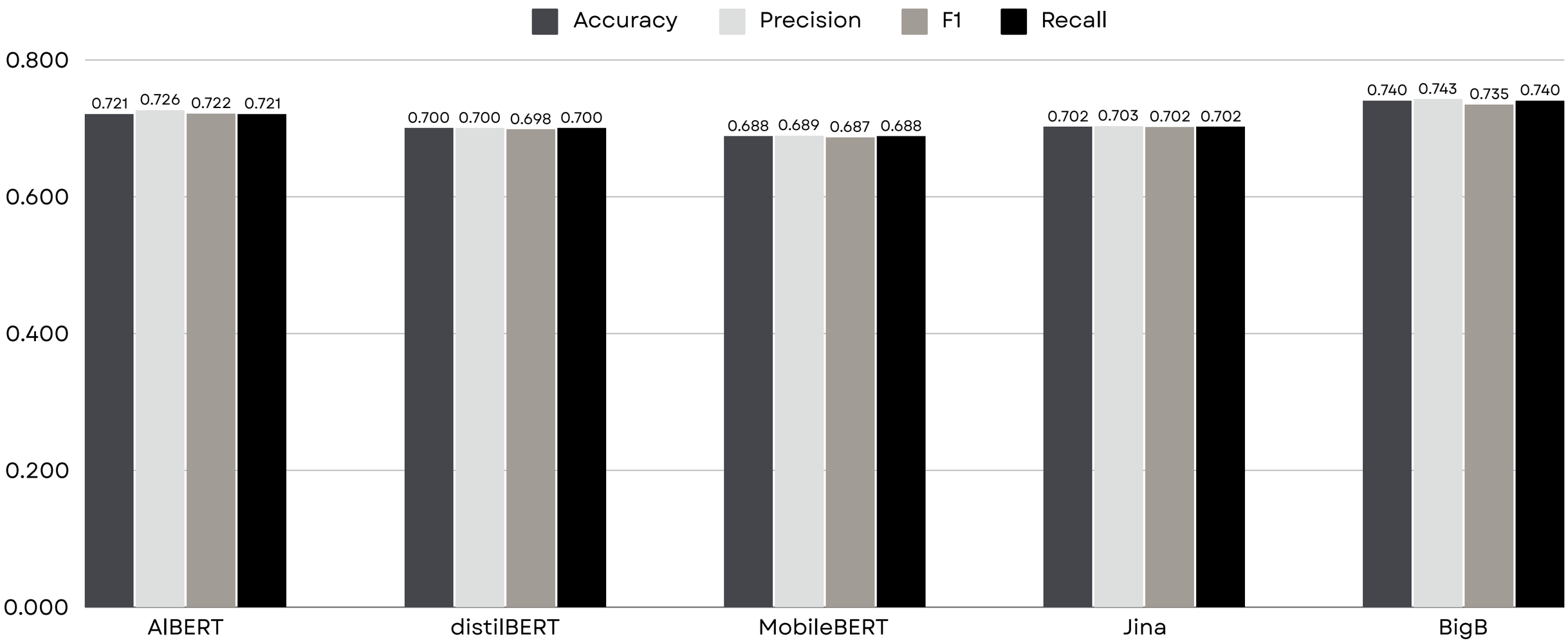
Model Comparisons for each task

EI SUBTASK



*denotes precision <0.4

NLI SUBTASK


Accuracy
Precision
F1
Recall
0.800
0.600
0.400
0.200
0.000
0.721
0.726
0.722
0.721
0.700
0.700
0.698
0.700
0.688
0.689
0.687
0.688
0.702
0.703
0.702
0.702
0.740
0.743
0.735
0.740
AlBERT
distilBERT
MobileBERT
Jina
BigB