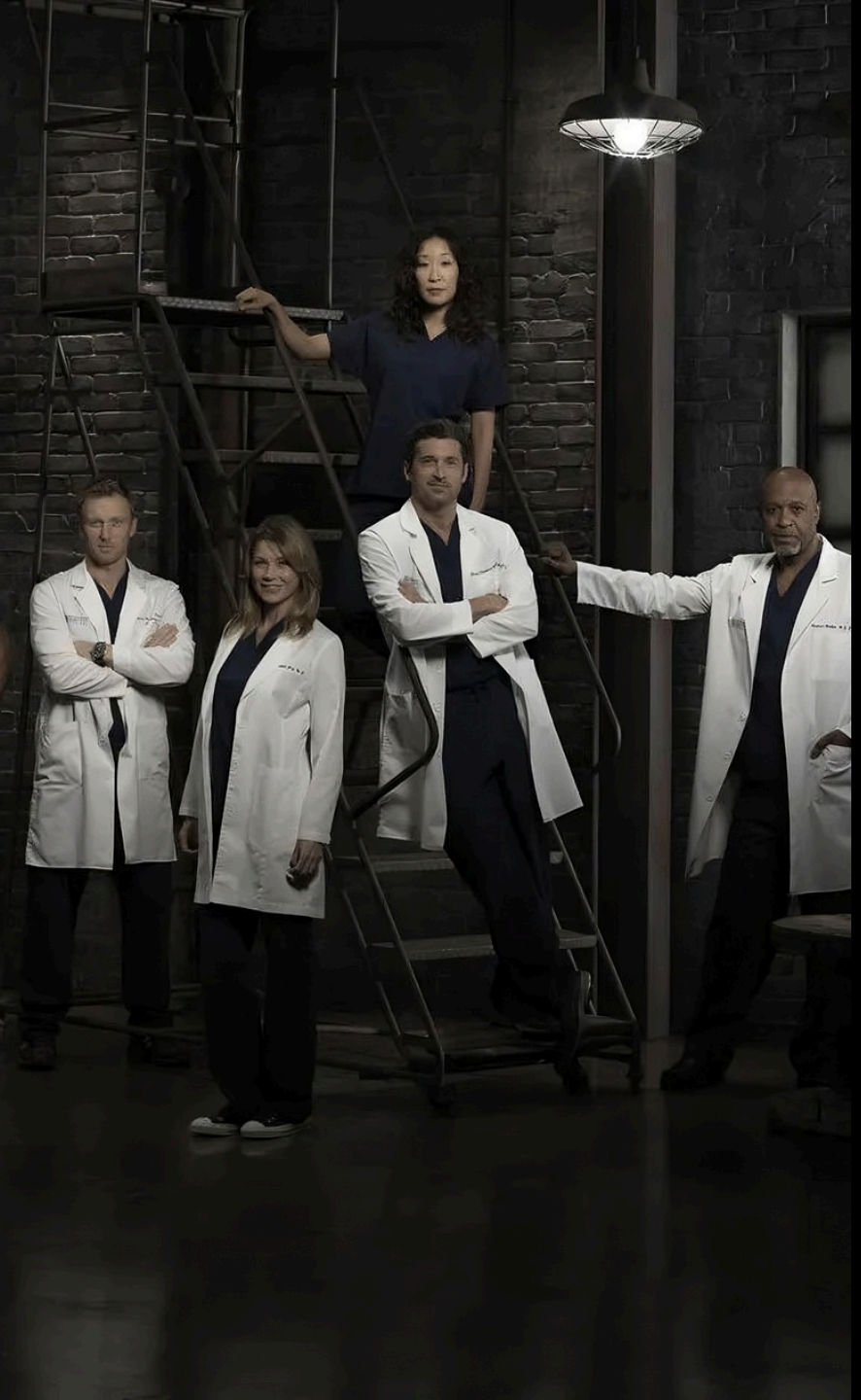


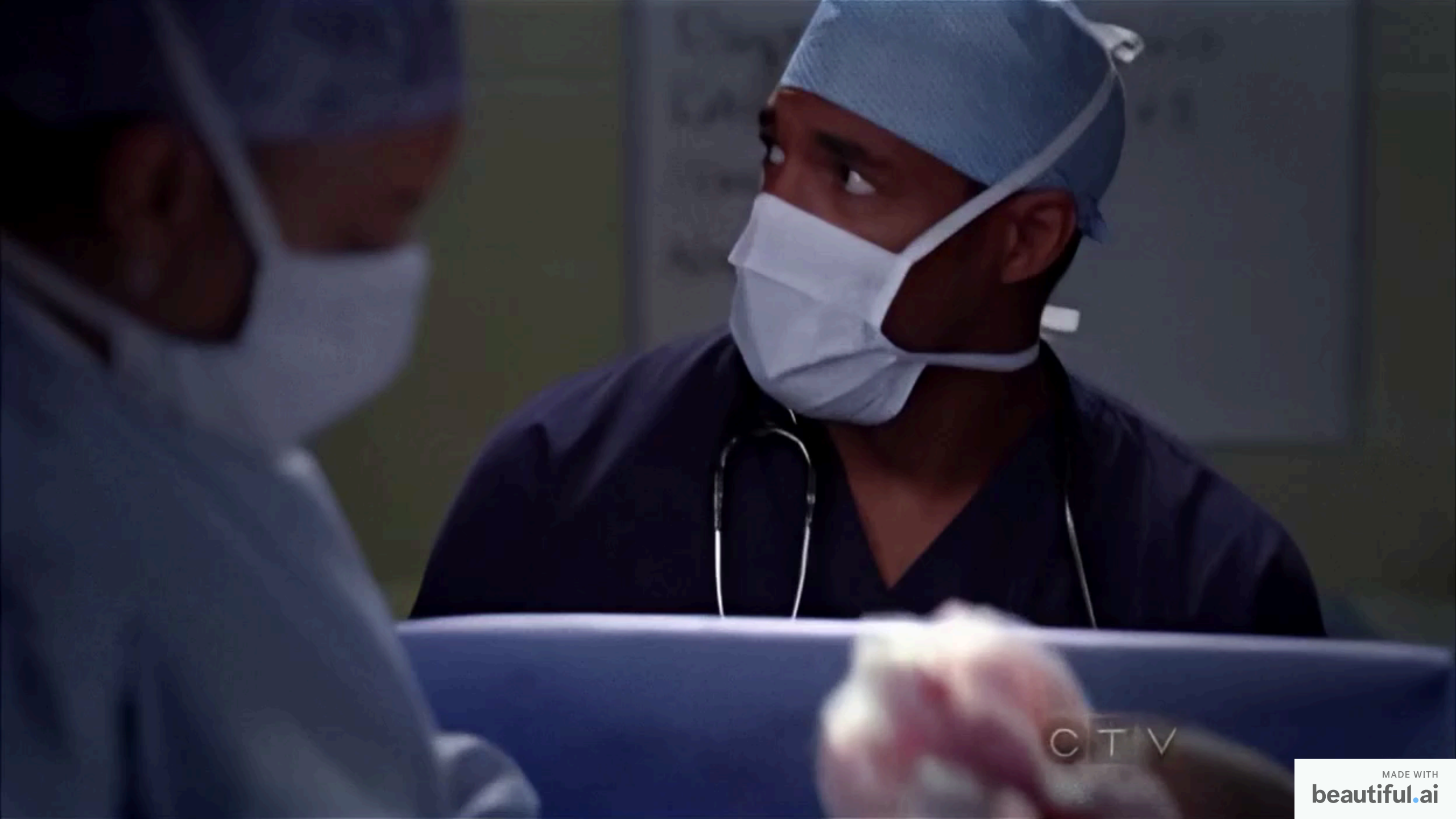
Toward the automatic identification of isotopies

Alice Fedotova, Alberto Barrón-Cedeño

Department of Interpreting and Translation, University of Bologna

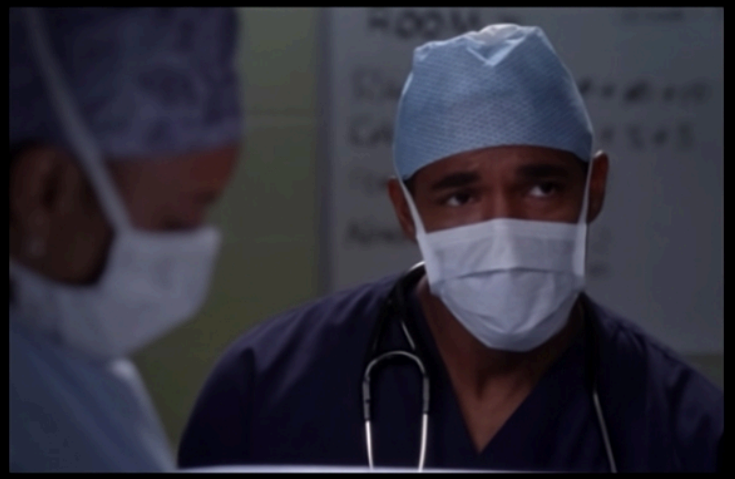
Outline

1. Task Definition
2. Isotopies
3. Research Questions
4. Methodology
5. Dataset
6. Models
7. Evaluation
8. Results and Discussion
9. Future Work



Task Definition

Classification: identifying medical drama narratives



"How was your lunch Dr. Grey?"

Model

Sentimental Plot

Professional Plot

Medical Case

Isotopies

The classes that the model learns to identify

Professional Plot

Sentimental Plot

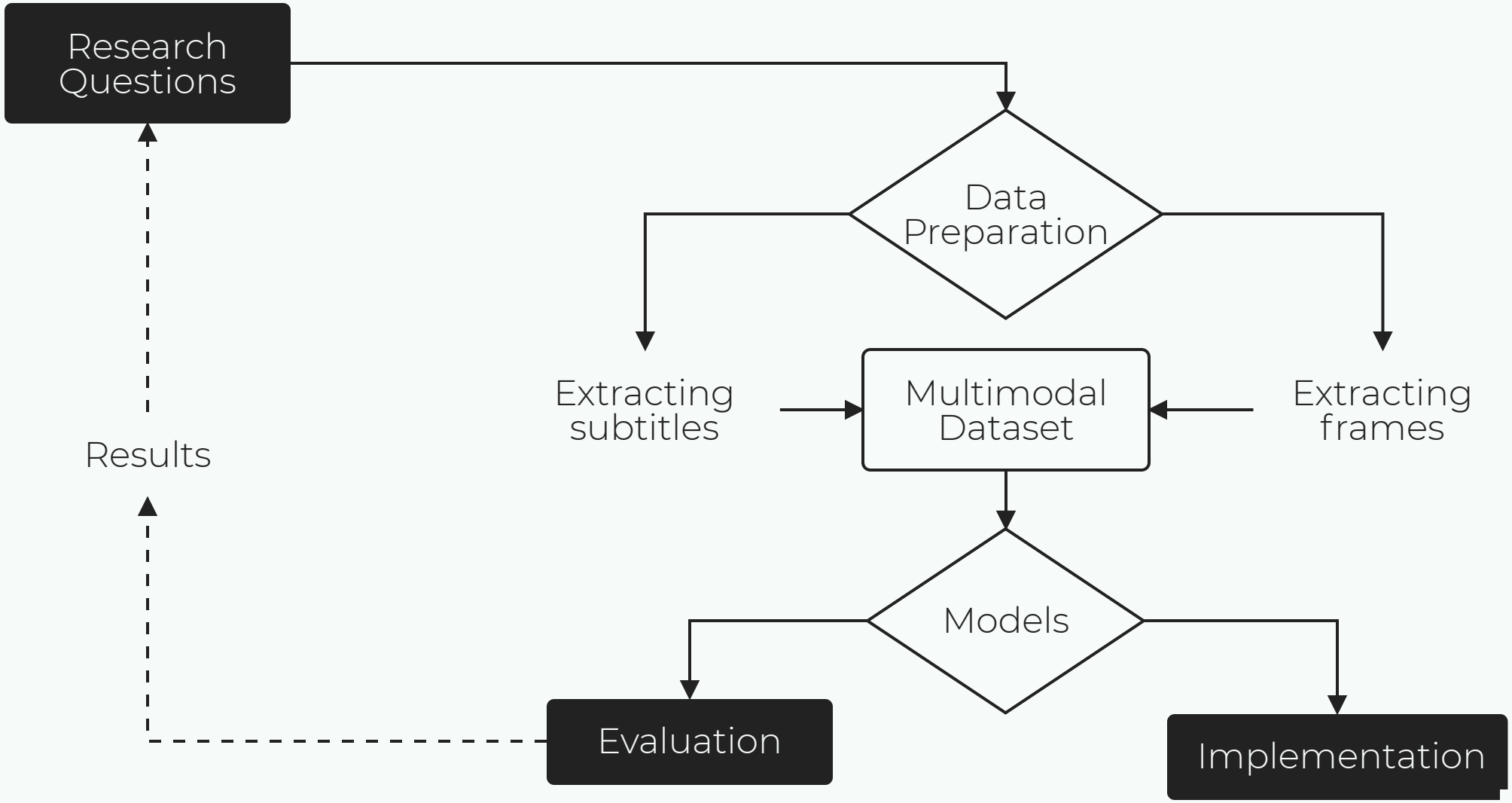
Medical Cases Plot

Rocchi and Pescatore (2022) - Modeling narrative features in TV series: coding and clustering analysis

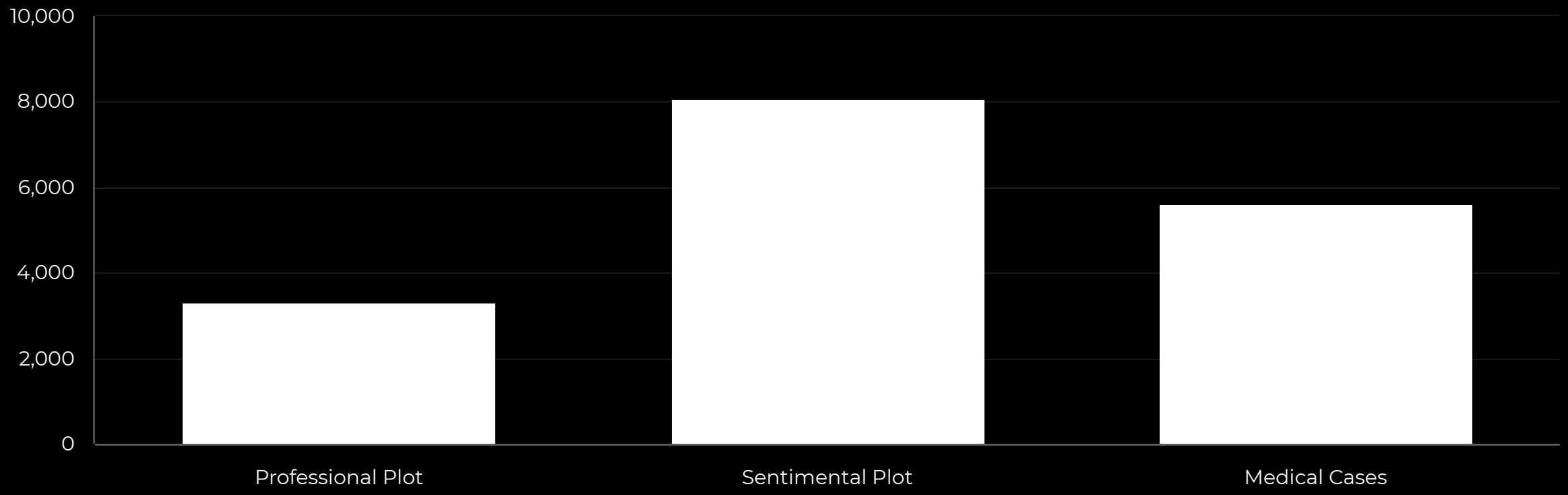
Research Questions

- **Research Question 1**
  Approach comparison (multiclass vs one-vs-the-rest)
- **Research Question 2**
  Modality comparison (text vs frames)
- **Research Question 3**
  Combining the modalities (does it help?)

Methodology



Dataset



The final dataset contains 16,989 instances from Grey's Anatomy.
The most represented class is SP, while the least represented is PP.

Models

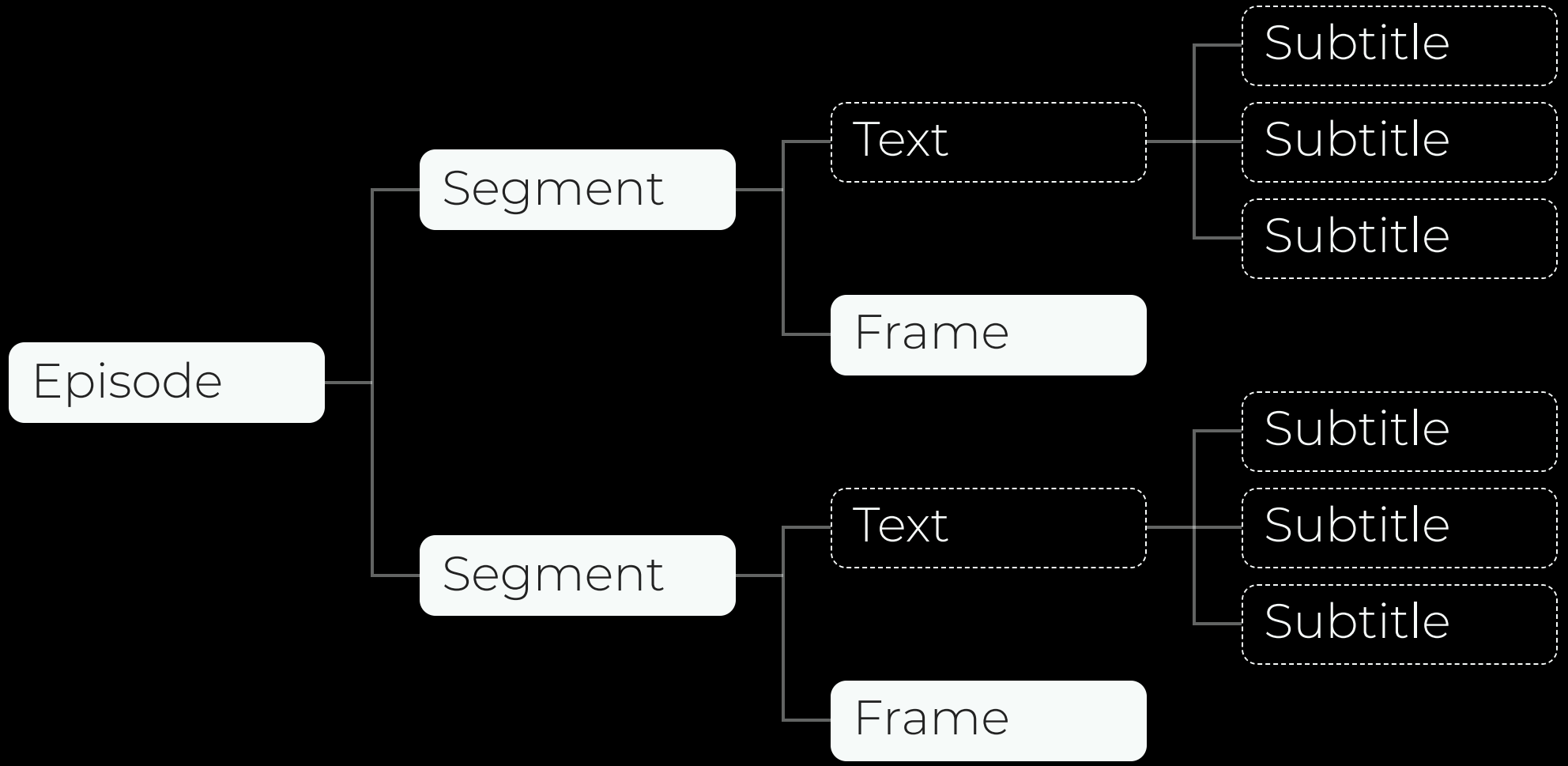
1. **Unimodal visual**
   Based on frames

2. **Unimodal textual**
   Based on subtitles

3. **Multimodal**
   Based on both

CLIP

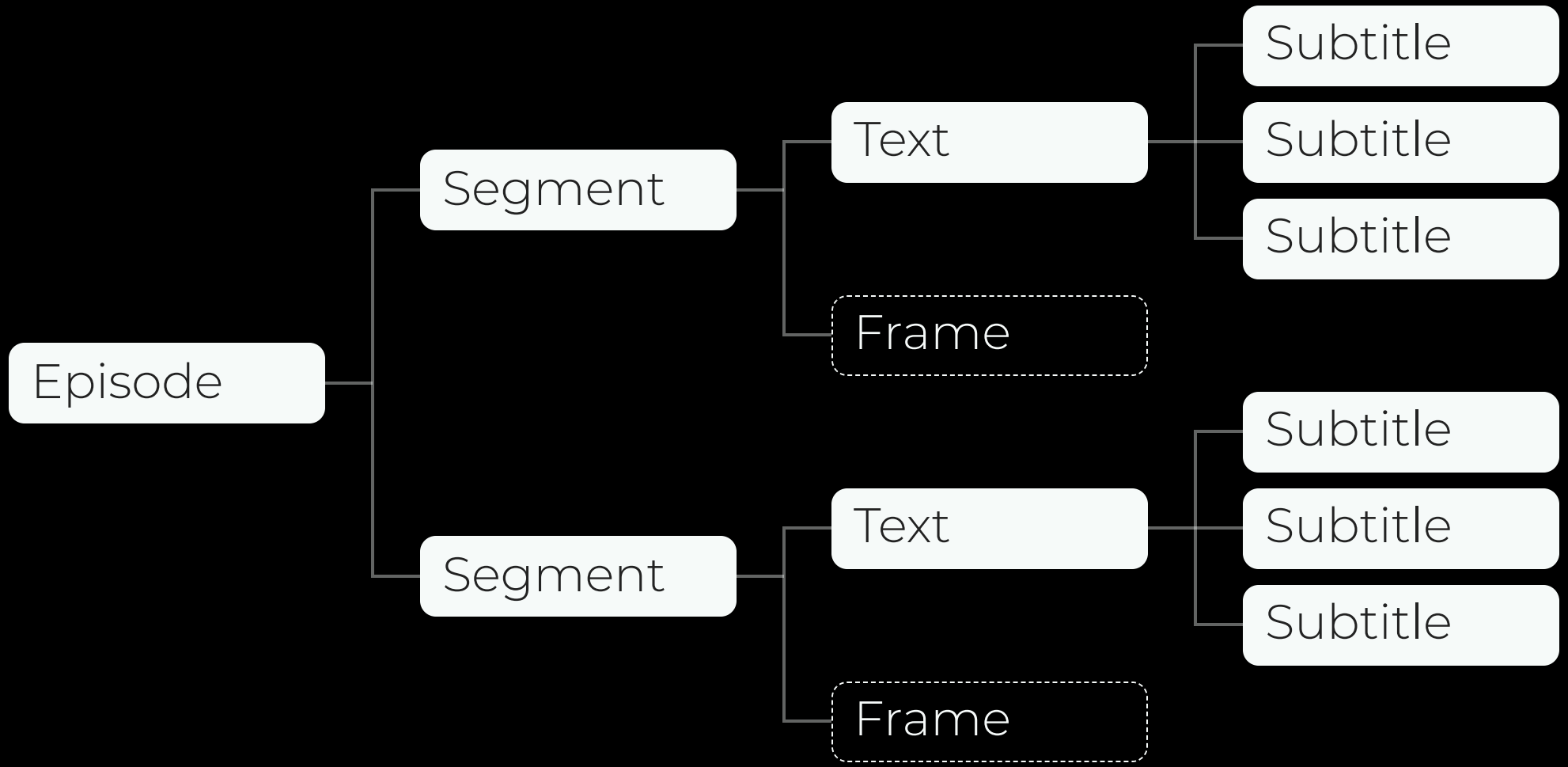
Based on the frames, the model learns to assign isotopies to segments



Radford et al., 2021

BERT

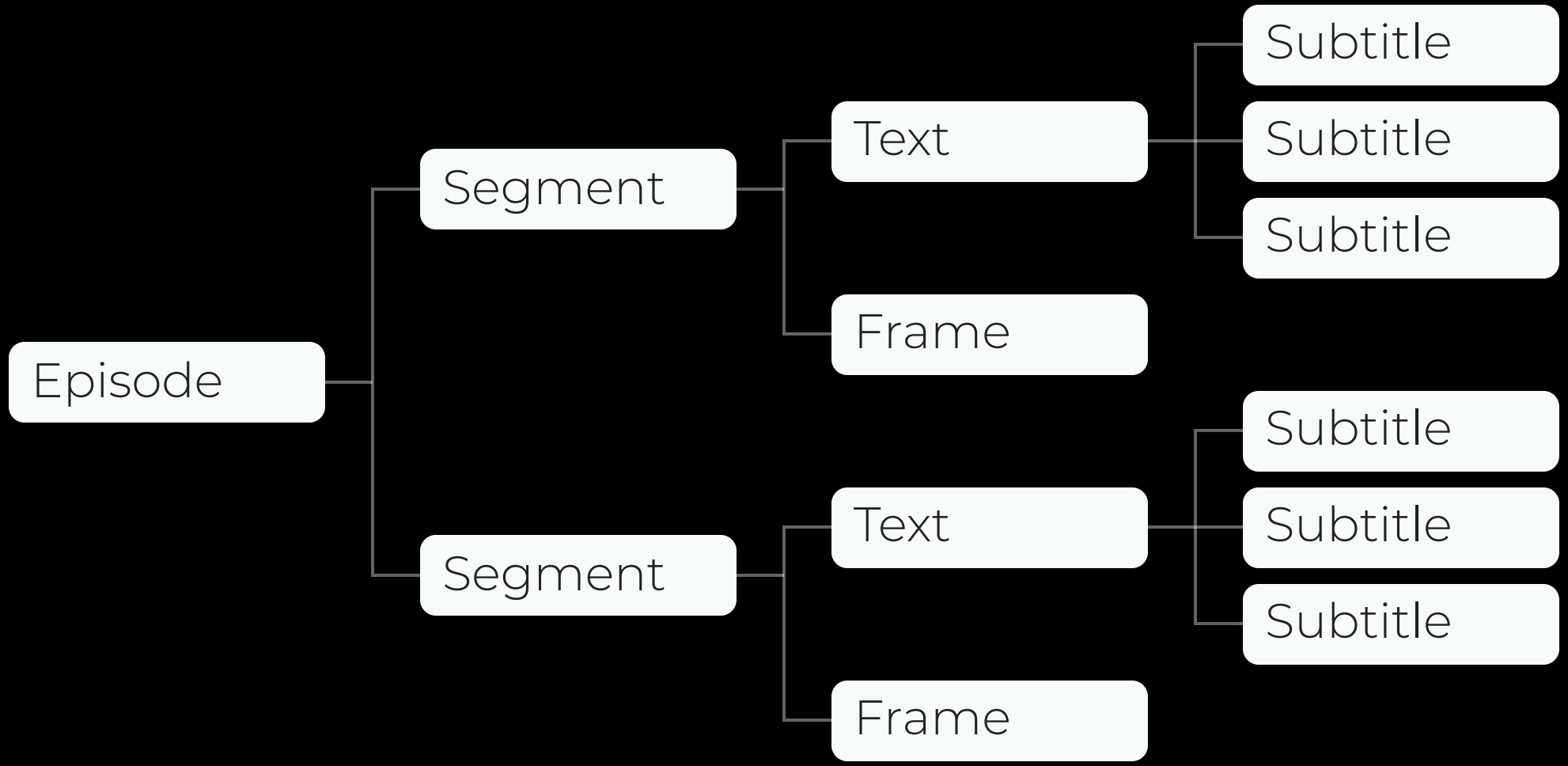
Based on the text, the model learns to assign isotopies to segments



Devlin et al., 2019

MMBT

Based on text and frames, the model learns to assign isotopies to segments



Kiela et al., 2020

F1 Score

The higher the better

Precision

Of all the instances identified as sentimental plots, how many were actually sentimental plots?

Recall

Of all the sentimental plots, how many did the model correctly identify as sentimental plots?

F1 Score

Considers both Precision and Recall. It is a value from 0 to 1.

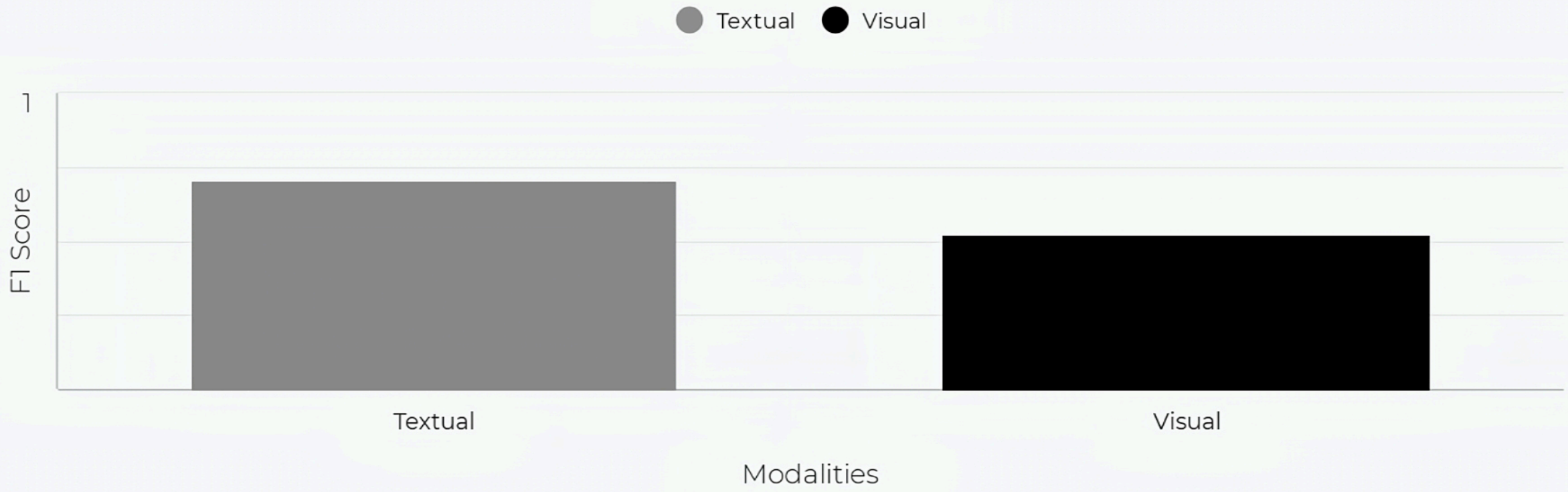
RQ1

Is it better to approach the task with a single multiclass model or a one-vs-the-rest approach?

F1 scores of the best configurations on test.

RQ2

Which modality is more informative: visual or textual?



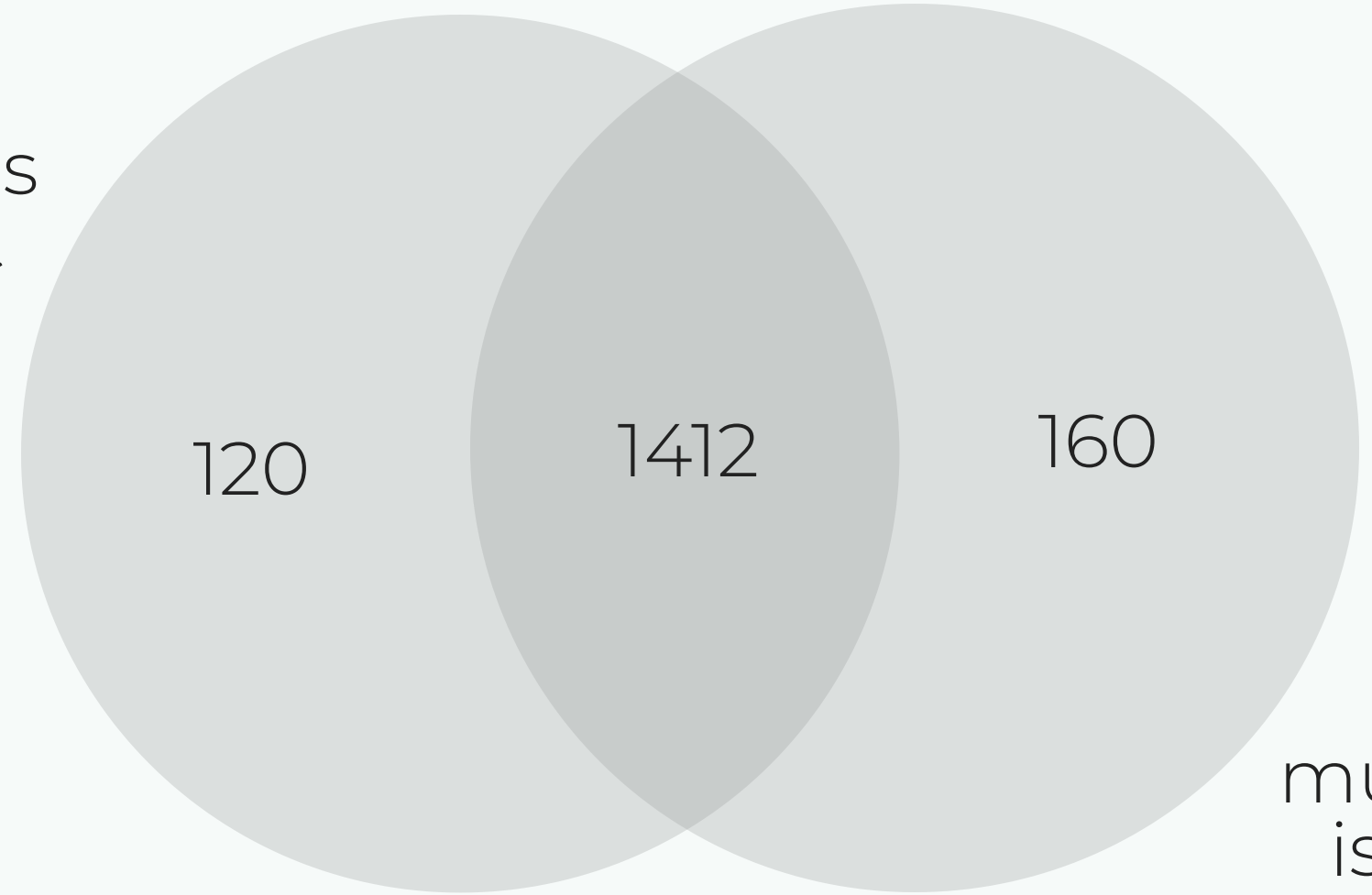
The highest F1 score (0.685) on test was obtained by the textual model.

RQ3

Does the inclusion of keyframes improve performance?

Highest F1 scores by model on test. Multimodal outperforms textual

Errors



More on errors

Conflicting visual and textual information

An example

Multimodal is wrong, textual is right

"... Dude, you're operating in the dark? On your wife?"

Multimodal Model

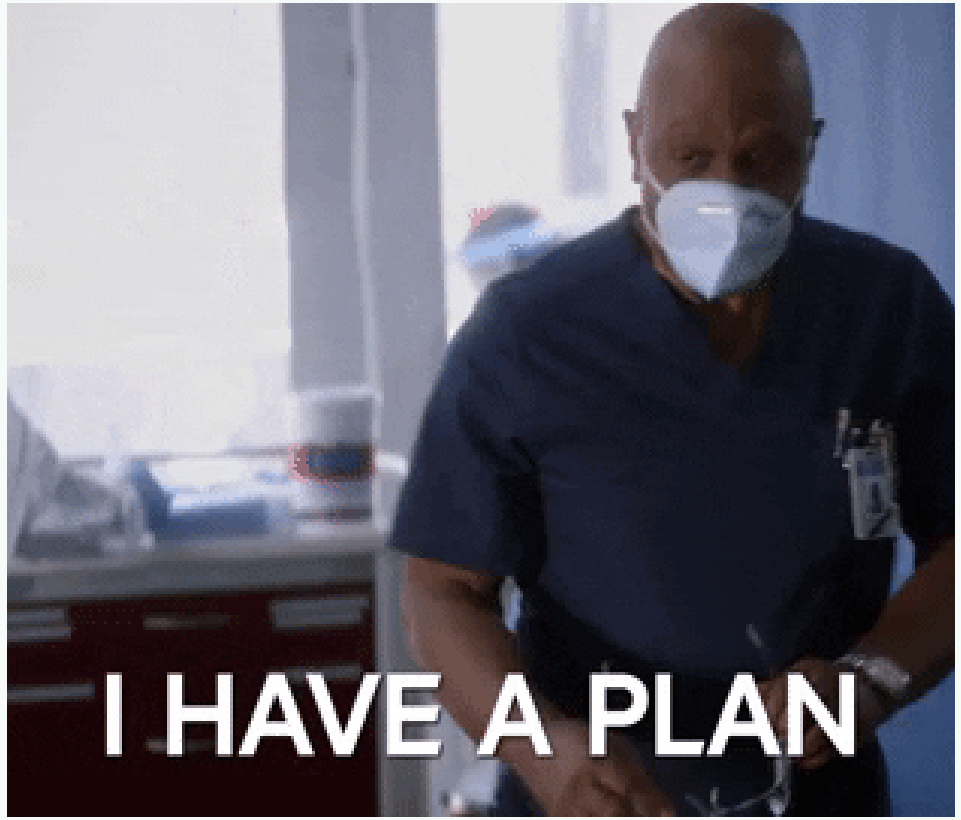
Sentimental Plot

Professional Plot

Medical Case

Future Work

1. Multilingual models
2. Considering more frames
3. Assessing generalization



Thank you for your attention!