Model	Dev	Test
XLNet + RoBERTa (LDL)	0.547	0.518
XLNet + BiLSTM-ELMo (Keyphrase)	0.538	0.532
XLNet + BiLSTM-ELMo (LDL)	**0.55**	**0.543**

Table 7: Performance of different ensemble models

sponds to a particular sentence belonging to a presentation slide in the original corpus. The development set results can be found in Table 8. The evaluation scheme used in this experiment uses the same Match$_m$ as described in the Evaluation Metric section but with m = $1, 2, 3, 4$ as used in **?**.

Model	Dev
XLNet	**0.758**
XLNet (LDL)	0.757
RoBERTa	0.743
RoBERTa (LDL)	0.745
BiLSTM-ELMo	0.751
BiLSTM-ELMo (LDL)	0.752

Table 8: Sentence-wise results on the Development set

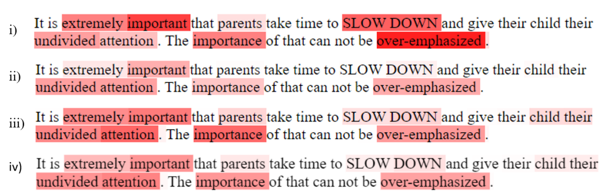

Figure 5: Emphasis Heatmaps i) Ground Truth ii) BiLSTM-ELMo iii) XLNet iv) Best Ensemble Model

Analysis

Length vs Performance

We wanted to understand how the performance of our models was affected by the length of the instances. Table 9 summarizes the performance of our best performing single model, i.e, XLNet on the development set divided into three sets, Short ($\leq$ 40 tokens, 80 samples), Medium (40 to 90 tokens, 262 samples), and Long (>90 tokens, 50 samples). As we can see, the model performance deteriorates with the increasing length of the instances.

	XLNet
Small ($\leq$ 40)	**0.648**
Medium (>40 and $\leq$ 90)	0.549
Large (>90)	0.42

Table 9: Average $Match_m$ for best performing XLNet model on different size of instances in the development set

Emphasis vs Parts of Speech

Table 10 shows POS (Parts of Speech) tags vs. average emphasis on the development dataset. We did this experiment to understand how our model predictions performed on each POS tag when compared to the actual human-annotated emphasis scores on the development set. We noticed that the original average emphasis scores were highest on Adjectives followed by Noun. On comparing our models, we found that XLNet was able to almost accurately predict the emphasis scores on Adjectives and Noun respectively, and BiLSTM-ELMo also had the highest predictions on Adjectives and Noun respectively. We also noticed that XLNet did a better job on predicting the emphasis score on different POS tags where the predictions were either very close to the human scores or marginally lesser. On the other hand, we noticed that BiLSTM-ELMo's predictions fell short by bigger margins when compared to XLNet and gave more emphasis to Adverbs than that in the development set.

POS	Count	Human	BiLSTM	XLNet
Noun	4719	0.169	0.134	0.168
Verb	1420	0.118	0.083	0.113
Adjectives	982	0.186	0.140	0.181
Det	634	0.062	0.029	0.042
Adverbs	347	0.111	0.068	0.103
Pronouns	165	0.040	0.068	0.022
Punct	2082	0.034	0.015	0.025

Table 10: POS tags vs. average emphasis on development dataset

Conclusion

In this paper, we present our approach to AAAI-CAD21 shared task: Predicting Emphasis in Presentation Slides. Our best submission gave us an average $Match_m$ of 0.518 placing us 3rd on the Evaluation phase leaderboard and an average $Match_m$ of 0.543 placing us 1st on the Post-Evaluation leaderboard at the time of writing the paper. Future work includes using a hierarchical approach to emphasis prediction as a sequence labeling task using both sentence-level (individual sentence in a slide) and slide-level representations of a word (**?**).

Acknowledgement

Rajiv Ratn Shah is partly supported by the Infosys Center for AI at IIIT Delhi. We also thank Sunny Dsouza and Gautam Maurya for their detailed and valuable feedback.

Neque tenetur asperiores repellendus maiores aspernatur molestiae placeat, distinctio quod tempore ab sit eos id repellat enim quis soluta tempora, quisquam praesentium similique ea error, voluptate blanditiis harum architecto aliquid recusandae omnis, nihil eum culpa dicta ut repudiandae facere dolorum obcaecati asperiores sequi est?Quod sapiente laboriosam libero maiores perspiciatis aliquam, omnis dolorem excepturi, esse possimus debitis dolore animi laudantium quaerat dicta excepturi mollitia nam ducimus?Porro minus maxime nisi consequatur ab necessitatibus nesciunt, odit harum commodi, laudantium quas tenetur inventore beatae aperiam laboriosam, quaerat dolor quam modi, quidem porro sequi aliquam