

**Le et al. reproduced results using BERT (huggingface):**

Model dimensions: **BERT base uncased, max sequence length: 128 tokens, batch size: 16**

Training was completed in **32:08** minutes and **3** epochs; train, dev and test set were same as within Le (2020)

Overall used samples: **99299**

Samples used for training: **79439**

Samples used for development: **9930**

Samples used for testing: **9930**

	Precision	Recall	F1-score	support
<i>BQ</i>	<b>0.96147</b>	<b>0.95147</b>	<b>0.95644</b>	<b>3750</b>
<i>BS</i>	<b>0.96901</b>	<b>0.97485</b>	<b>0.97192</b>	<b>11291</b>
<i>O</i>	<b>0.99803</b>	<b>0.99774</b>	<b>0.99788</b>	<b>100447</b>
<i>accuracy</i>			<b>0.99400</b>	<b>115488</b>
<i>Macro avg</i>	<b>0.97617</b>	<b>0.97468</b>	<b>0.97541</b>	<b>115488</b>
<i>Weighted avg</i>	<b>0.99400</b>	<b>0.99400</b>	<b>0.99400</b>	<b>115488</b>