Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220609

Deep Learning-Based Brain Hemorrhage Detection in CT Reports

Gıyaseddin BAYRAK [a,1], Muhammed Şakir TOPRAK [b,c] Murat Can GANİZ [a]
Halife KODAZ [c] and Ural KOÇ [b]
[a] *Computer Engineering Department, Marmara University, Turkey*
[b] *Ministry of Health, Turkey*
[c] *Computer Engineering Department, Konya Technical University, Turkey*

Abstract. Radiology reports can potentially be used to detect critical cases that need immediate attention from physicians. We focus on detecting Brain Hemorrhage from Computed Tomography (CT) reports. We train a deep learning classifier and observe the effect of using different pre-trained word representations along with domain-specific fine-tuning. We have several contributions. Firstly, we report the results of a large-scale classification model for brain hemorrhage detection from Turkish radiology reports. Second, we show the effect of fine-tuning pre-trained language models using domain-specific data on the performance. We conclude that deep learning models can be used for detecting brain Hemorrhage with reasonable accuracy and fine-tuning language models using domain-specific data to improve classification performance.

Keywords. NLP, Deep Learning, Brain Hemorrhage, Radiology

1. Introduction and Methodology

There are several studies that develop classification models for radiology reports. For example, [1] studied Epilepsy classification using bi-LSTM on a small dataset of radiology reports from MRI. Our contributions can be summarized as follows: 1) An implementation for the critical non-traumatic hemorrhage detection from radiology reports. 2) A comparison between the baseline pre-trained FastText [2] and BERT [3] language models and task-specific fine-tuned variations of them. 3) One of the first studies to train deep learning models on a large textual dataset consist of Turkish radiology reports. [2]

The brain CT reports are labeled using the 10_{th} version of the International Classification of Diseases (ICD-10) diagnostic codes. There are two ICD-10 codes assigned to each report, the first for preliminary diagnosis, assigned before the diagnostic imaging, and the second for the final diagnosis after examining the images and radiology report. We use about 100,000 brain CT reports for patients with a preliminary diagnosis of hemorrhage indicated with I60, I61, I62. The reports whose preliminary and final di-

[1] Corresponding Author: Istanbul, Turkey; E-mail: giyaseddinalfarkh@marun.edu.tr.
[2] The use of data is approved under ethics vote number E.Kurul-E1-22-2326 by the ethical committee of the Ministry of Health. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the Ministry of Health of the Government of Turkey.

Pre-trained Embedding Layer	**Precision**$_{Positive}$	**Recall**$_{Positive}$	**Macro F-1**
1) $FastText_{common\ crawl}$[3]	$79.06_{\pm 0.71}$	$21.81_{\pm 0.98}$	$55.17_{\pm 0.64}$
- $FastText_{common\ crawl\ (frozen)}$	$74.82_{\pm 3.42}$	$23.98_{\pm 2.18}$	$55.98_{\pm 0.94}$
2) $FastText_{pre-trained\ on\ 190k\ unsupervised\ reports}$	$80.85_{\pm 1.09}$	$21.25_{\pm 0.42}$	$54.98_{\pm 0.32}$
- $FastText_{pre-trained\ on\ 190k\ unsupervised\ reports\ (frozen)}$	$78.58_{\pm 2.11}$	$23.03_{\pm 1.10}$	$55.92_{\pm 0.52}$
3) $Randomly\ initialized\ embeddings_{No\ pre-trained\ weights}$	$68.16_{\pm 1.56}$	$50.35_{\pm 4.21}$	$66.89_{\pm 0.95}$
4) $BERT_{Base}$[4]	$72.92_{\pm 2.50}$	$47.49_{\pm 2.60}$	$67.53_{\pm 0.41}$
5) $BERT_{Fine-tuned\ on\ trainingreports}$	$71.93_{\pm 2.82}$	$57.76_{\pm 3.88}$	$71.07_{\pm 0.55}$
6) $BERT_{Fine-tuned\ on\ 190k\ unsupervised\ reports}$	$74.39_{\pm 3.98}$	$58.66_{\pm 6.44}$	$\mathbf{72.21}_{\pm 0.97}$

Table 1. Precision, Recall and F1 scores of bi-LSTM classifier with different word representations.

agnoses codes match are labeled as positive (15697), the rest as negative (21819). Our training, validation, test split ratios are 64, 16, and 20, respectively. For pre-training and fine-tuning of language models (LM) we use 190,000 brain and thorax CT reports. We use the same deep learning classifier [4] bi-LSTM with fixed hyper-parameters with different word representations [5] to see the effect of pre-trained models and fine-tuning. Bi-LSTM is trained for 4 epochs using ADAM optimizer with a learning rate of 0.001.

2. Results and Conclusion

We show the effect of using different word representations in Table 1. As can be seen in the table, the choice of word representation, e.g. fastText or BERT and its fine-tuning has a drastic effect on the performance of the classifier. As expected, BERT contextual embeddings work better than static embeddings of fastText. Fine-tuning BERT pre-trained model with smaller but labeled task-specific data makes a difference. The difference is most visible when we fine-tune BERT with larger amounts of unsupervised domain-specific data.

We conclude that deep learning models can be used for detecting Brain Hemorrhage from radiology reports with reasonable performance ($\simeq 72\%$ F1) and fine-tuning pre-trained language models such as BERT using domain-specific data to improve classification performance. In the future, we plan to develop a highly accurate and explainable real-world classification system to detect critical brain hemorrhage.

References

[1] Bayrak S, Yucel E, Takci H. Epilepsy Radiology Reports Classification Using Deep Learning Networks. COMPUTERS, MATERIALS AND CONTINUA : Tech Science Press. 2022;70(2):3589-607.

[2] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics. 2017;5:135-46.

[3] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.

[4] Bustos A, Pertusa A, Salinas JM, de la Iglesia-Vayá M. Padchest: A large chest x-ray image dataset with multi-label annotated reports. Medical image analysis. 2020;66:101797.

[5] Drozdov I, Forbes D, Szubert B, Hall M, Carlin C, Lowe DJ. Supervised and unsupervised language modelling in Chest X-Ray radiological reports. Plos one. 2020;15(3):e0229963.

[3] The publicly available fastText model https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.tr.300.bin.gz

[4] The publicly available model https://huggingface.co/dbmdz/bert-base-turkish-uncased