Australian e-Health Research Centre

Developing Robust Clinical Text Deep Learning Models – a "Painless" Approach

Yutong WU | Postdoctoral Fellow

James A HUGHES, Anna-Lisa LYRSTEDT, Sarah HAZELWOOD, Nathan J BROWN, Lee JONES, Clint DOUGLAS, Rajeev JARUGULA, Kevin CHU, Anthony NGUYEN

Australia's National Science Agency

Background

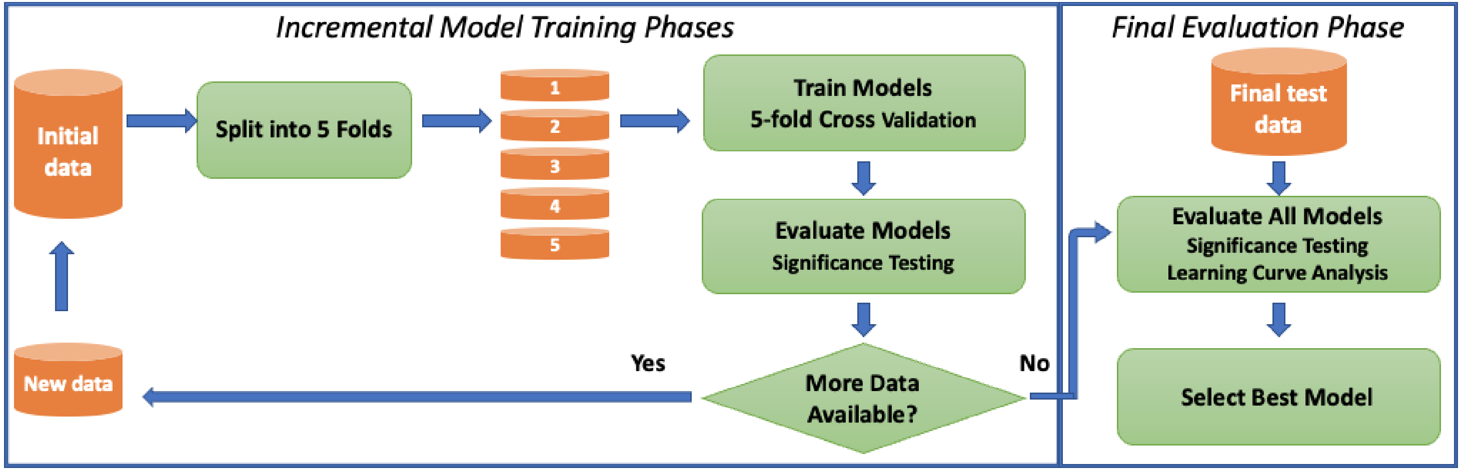
- Deep learning relies on large-scale labeled training data
- Labelling clinical data is costly and labour-intensive, requiring highly trained annotators
- Data scarcity hinders deep learning in health/clinical applications

Objectives

- Develop robust clinical text deep learning models with limited labeled data
- Determine sufficient dataset sizes for training-validation-testing
- Understand trade-offs between model performance and resource cost (training set size, learning time, model size)

Methods

A ***generalised incremental multiphase methodological framework*** was used to develop and robustly assess clinical text classifiers.



Methods

Learning curve analysis was conducted on diverse deep learning model architectures to:

- Generalize findings
- Identify performance-resource trade-offs
    - Due to varying parameters, computational costs, and domain specificities

Data description – *Pain dataset*

- 10,000 randomly selected records
  - routinely-collected clinical information from adult patients at an Australian hospital ED between March 2018 and February 2021
- Two free-text data fields, completed by ED nurses when patient arrives
  - “presenting problem” and “nurse assessment”
- Annotated as “Pain” or “No Pain” by two ED nurses. Third nurse for adjudication if disagreement.
  - 5520 as “Pain”: patient presented with signs or symptoms of pain
  - 4480 as “No-Pain”
  - 957 records required adjudication

Problem formulation

- A **binary text classification** problem
  - Use concatenated triage nurse assessment (two free-text fields) to predict the pain outcome ("Pain" or "No-Pain")
- Research ethics was obtained from the Human Research Ethics Committee of the Royal Brisbane and Women's Hospital (LNR/2021/QRBW/72976) and the Queensland University of Technology (109147)

Deep learning models

Popular text classification models ranging from simple to complex deep learning architectures (including transformers).

- Fine-tuned ***BERT*** and ***Distilled-BERT***: large pretrained general domain Transformers[1] and its distilled version[2] with reduced parameter space
- Fine-tuned *domain-specific* ***BlueBERT*** and ***Distilled-BlueBERT***: large biomedical domain variant of BERT[3] and its distilled version with reduced parameter space
- ***AttRNN-BERT*** and ***AttRNN-BlueBERT***: attention-based RNN models utilizing BERT/BlueBERT as word embeddings with RNN layer[4, 5]

Australian e-Health
Research Centre

Results

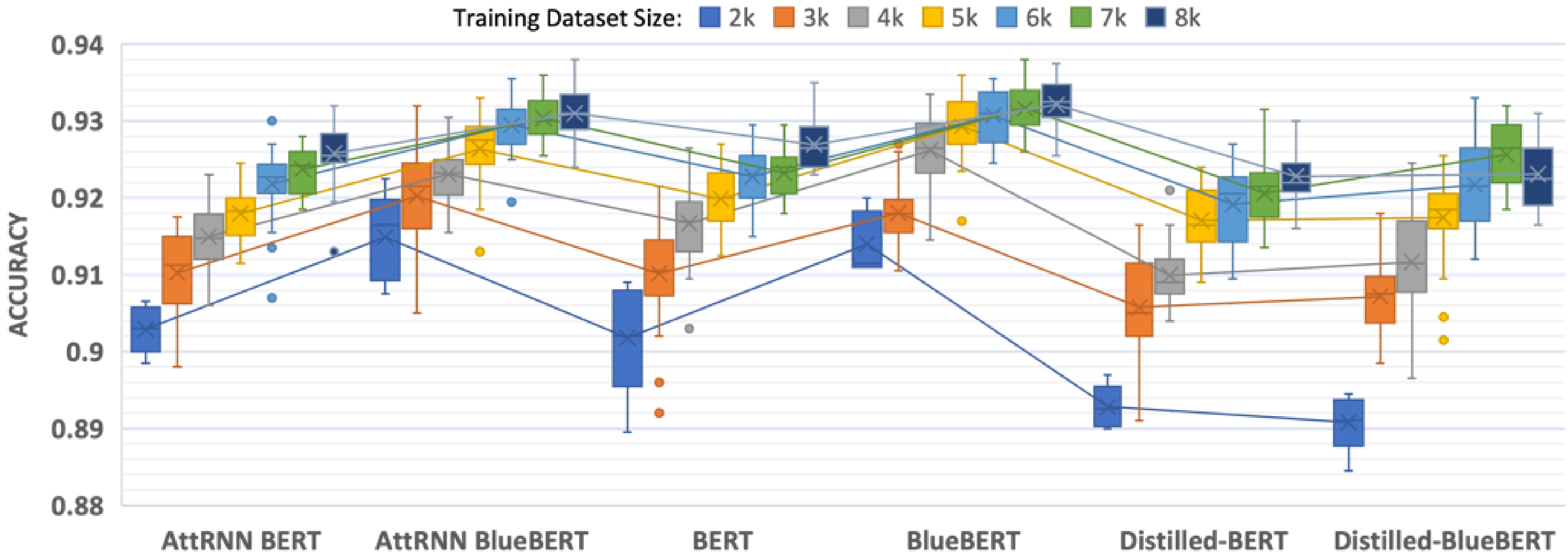


Figure 2. Learning curves of deep learning models evaluated on the final evaluation dataset.

Results

- Health domain-specific pre-trained models showed superior classification performances compared to general language models (p-value < 0.05)
- Distilled models, although not as competitive as uncompressed versions (p-value < 0.05), offer utility in low-resource settings by achieving good performance with much smaller parameter sets

Conclusions

- Incremental multiphase framework effectively generates learning curves to capture performance trends and variability across models and dataset sizes
- Significance testing quantifies robustness and identifies ‘optimal’ dataset sizes
- Results confirm importance of domain-specific pre-training and distilled models in low-resource settings
- Framework is flexible, can adapt to different learning methods that require less supervision and can continuously update the models with new labeled data.

References

1. Banko M, Brill E. Scaling to very very large corpora for natural language disambiguation. ACL 2001 Jul (pp. 26-33).
2. Mukherjee S, Awadallah A. XtremeDistil: Multi-stage distillation for massive multilingual models. arXiv preprint arXiv:2004.05686. 2020 Apr 12.
3. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. arXiv preprint arXiv:1906.05474. 2019 Jun 13.
4. Vu T, Nguyen A, Brown N, Hughes J. Identifying patients with pain in emergency departments using conventional machine learning and deep learning. ALTA 2019 (pp. 111-119).
5. Hughes J, Brown N, Vu T, Nguyen A. Towards automatic and interpretable assignments of patients presenting with pain to the emergency department. Digit. Health Inst. Summit 2020, 2021 Apr;*276*, p.20.

Thank you

Health Data Semantics & Interoperability
Dr Yutong Wu
yutong.wu@csiro.au