Towards Automatic and Interpretable Assignments of Patients Presenting with Pain to the Emergency Department

JA HUGHES [a,b,1], NJ BROWN [a,c], Thanh VU [d] and Anthony NGUYEN [d]

[a] *Emergency and Trauma Centre, Royal Brisbane and Women's Hospital, Brisbane, Australia*
[b] *School of Nursing, Queensland University of Technology, Brisbane, Australia*
[c] *Faculty of Medicine, University of Queensland, Brisbane, Australia*
[d] *Australian e-Health Research Centre, CSIRO, Brisbane, Australia*

Introduction. Pain is the most common symptom that patients present with to the emergency department. It is hard to identify patients who have presented in pain to the emergency department when compliance with structured pain assessment is low. An ability to identify patients presenting in pain allows further investigation of the quality of care provided. **Background.** Machine and deep learning techniques are commonly used for text analysis in healthcare. Applications such as the classification of diagnosis and unplanned readmissions from textual medical records have previously been described. In other work, conventional and deep-learning techniques have demonstrated high performance in identifying patients presenting to the emergency department in pain. However, these models have lacked interpretability. **Methods.** This paper proposes the use of machine learning techniques to identify patients who present in pain based upon their initial assessment using interpretable deep learning models. **Results.** The interpretable deep learning model of pain identification was shown to have more accuracy and precision than other machine and deep learning techniques. This technique has significant application to large datasets for the identification of the quality of care and real-time identification of patients presenting in pain to improve their care.

Keywords. pain, Emergency Department, machine learning, deep learning, nursing assessment

1. Introduction

Pain is the most common symptom that patients experience when they present to the emergency department (ED) [1], with between 65% [2] and 78% [3] of all presentations experiencing pain. In the ED, pain assessment and treatment have been reported as poor, leading to increased wait times and unnecessary symptom burden and suffering [4; 5]. Identification of patients who present with pain has long been a pillar of pain care in the ED but is poorly completed. Pain is a subjective experience and is best described and identified by the person experiencing it [6]. Clinical staff have previously been shown to be poor predictors of the presence and severity of pain [7; 8], with the patient's self-

---

[1] Corresponding Author, Dr James Hughes, Emergency and Trauma Centre, Royal Brisbane and Women's Hospital, Butterfield Street Herston; E-mail: james.hughes@health.qld.gov.au.

report of pain the most reliable indicator. In the absence of documented pain assessment, identifying patients who may have presented in pain can be difficult, and therefore alternative methods for identifying patients in pain are needed for; identifying prevalence of pain on presentation, assessing the outcomes of quality improvement activities and for use in real time to prompt clinicians to assess and treat patients presenting in pain. Others have previously described a method of manually identifying patients in pain using free text triage nursing assessments [9] that is not reliant on severity scoring, however this methodology at a large scale would be time consuming, laborious and impractical.

Machine learning is widely used in healthcare for predictive tasks such as, cancer staging from pathology reports [10], diagnosis from medical records [11; 12], and mortality and unplanned readmissions [13]. The increasing availability of large datasets collected by electronic medical records allows machine learning to be used to improve the quality of care provided to patients [10-12; 14]. To overcome the limitations of manual identification, Vu et al. [20] proposed to use machine learning including both conventional and deep learning models [15-19] to learn patterns in text that identify patients presenting with pain to the ED. Although the proposed machine learning models achieved high performances on the task, they were "black boxes" and do not explain "*how the model makes its decisions,*" which is a fundamental question in healthcare analytics. This paper proposes a interpretable deep learning model to handle the problem.

2. Methods

This study aims to formally introduce and test a method of identification of patients presenting to the ED in pain using conventional machine learning and deep learning techniques. To achieve this aim, it is herein described the construction of a dataset of ED patients from a large inner-city adult ED to evaluate the proposed machine learning models. In these models, each patient is assigned as either "Pain" denoting they presented with pain or "No Pain" denoting they arrived without pain as a presenting symptom.

2.1. Task Description

The task described in this work is a binary description of the presence of pain on arrival to the ED, either the patient has "Pain" or "No Pain". To achieve this task, unstructured free text of the nursing assessment and presenting problem entered into the electronic medical record at triage is used. Table 1 demonstrates examples of the unstructured free text that indicated both the presence of "Pain" and "No Pain". As can be seen, by these examples, the task is complicated by shorthand notation, abbreviations and typographical errors.

Table 1. Examples of unstructured free-text from the presenting problem and nursing assessment fields of the electronic medical record. Highlighted terms are indicative of pain.

"Presenting Problem" and "Nursing Assessment"	Class
2/24 frontal headache/ photophobia/ lower l) back urinary incontinence	"Pain"
? Seizure activity// not on meds// elevated post drug use hx of drug use	"No Pain"

2.2. Dataset

The ED data manager extracted the dataset comprised of a random selection of 2000 patients presenting to an inner-city Australian ED between August and October 2018. A medical student under the supervision of a senior clinician/researcher assigned each patient either a "Pain" or "No Pain" label based on the information contained in the presenting problem and nursing assessment free text fields. The student searched the free text for keywords indicating pain, or for a pain intensity assessment ("xx/10") or descriptor ("mild pain"), and assigned a "Pain" or "No Pain" label to each case accordingly. This dataset was then split into three datasets (Training, Development and Test) as per Table 2 below.

Table 2. Basic Dataset Statistics

Dataset	Patients	Pain	No-Pain
Training	1200	574 (48%)	626 (52%)
Development	400	171 (43%)	229 (57%)
Test	400	193 (48%)	207 (52%)
Total	2000	938 (47%)	1062 (53%)

2.3. Interpretable Deep Learning Model

The interpretable deep learning model builds upon performant artificial recurrent neural networks (RNN) applied on the same dataset [20]. In contrast to Vu et al. [20], an attention layer is added to help the model attend to important information (i.e., input words). The attention weights given by the layer for each input word can be used to help explain what the model did to produce the output label. In particular, more important words contributing to a label would be associated with higher attention weights. Figure 1 illustrates our AttRNN model architecture.
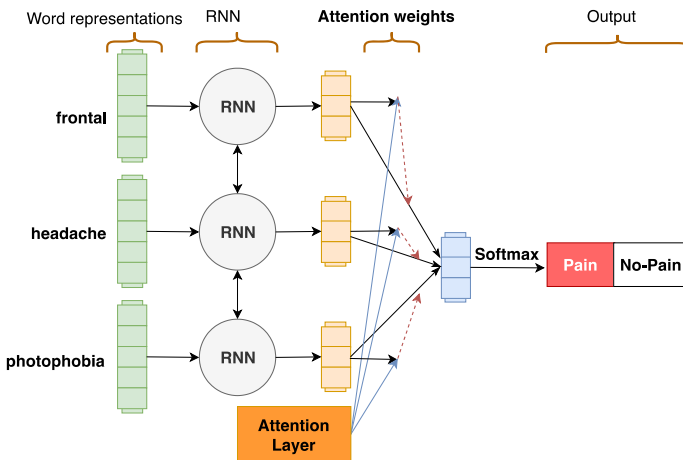

Figure 1. Attention-based Recurrent Neural Network model architecture (AttRNN)

The input to the model consists of a vector representation of words from the electronic medical record of a patient. The vector representations were learnt in advance

on a large-scale clinical dataset using a word representation model. *RNN* takes the vector representation of each word and produces an output vector for that word. The attention layer produces *attention weights* for each word. The final vector was generated using the *RNN* output vectors and the attention weights. The final vector was then used to predict the class of the input record (i.e., either "Pain" or "No-Pain") as the output of the model. In Figure 1, the model classifies a "Pain" label for the "… frontal headache photophobia …" input text. The model was learnt and finetuned on the Training and Development set, respectively. The model effectiveness will be reported on the unseen Test set.

3. Results

The performance of the interpretable deep learning model, AttRNN, was compared with other models proposed by Vu et al. [20], RULE - a rule-based model, Support Vector Machine (SVM) - a conventional machine learning model and Recurrent Neural Network (RNN) – a deep learning model. Table 3 shows the experimental results of the models on the test set. Note that AttRNN is an extension of RNN with an additional attention layer which helps explain the model output. In particular, we can see that the results produced by AttRNN are competitive with the state-of-the-art results produced by RNN, a deep learning model. More importantly, the produced attention weights of AttRNN can be used to visualise the importance of each word within the input text making the model interpretable. Table 4 shows examples of using attention weights produced by AttRNN to help interpret the results. In particular, "*headache photophobia*" and "*pain behind l ear*" are highlighted as more relevant to the "pain" label. It is worth noting that this interpretability feature is not available with RNN.

Table 3. Experimental results (%) on the test set. ∗ indicates that the performance difference between the machine learning models and the RULE model is significant at the significance level α of 0.1 using the Approximate Randomisation test [3, 5]. + indicates that the results are taken from Vu et al. [20].

Model	Accuracy (%)	Macro-averaged		
		Precision (%)	Recall (%)	F_1 (%)
RULE[+]	84.75	84.87	84.63	84.69
SVM[+]	88.00	88.13*	87.90*	87.96*
RNN[+]	**91.00***	91.21*	**90.88***	**90.96***
AttRNN	**91.00***	**91.43***	90.83*	90.94*

Table 4. AttRNN generated attention weights for interpreting the output.

Output	Input Interpretation
Pain	2 24 frontal headache photophobia lower l back urinary incontinence
Pain	Aloc and seizure activity post 2 unit blood donation today pain behind l ear

4. Discussion

The immediate application of the research is to provide interpretable machine learning assistance to identify patients presenting in pain over a very large, statewide dataset. This will allow the authors to answer the initial question of pain prevalence on presentation to the ED in Queensland and lead to the ability to explore the assessment and treatment of pain over a large population. Future applications include the potential real-time clinical application, such as a smart support assistant to help improve the quality of triage related to presentations that involve or are likely to involve pain. The gold standard in pain care is still universal assessment, clinicians at the bedside must assess and document a patient's pain. Tools such as the one described in this work will hopefully assist this in the longer term.

This paper proposed an interpretable deep learning model for the task of identifying patients who presented to EDs with pain. Experimental results on a 2,000 ED patient annotated dataset showed that not only this model performed well on this task with the highest accuracy and macro-averaged F_1 score of 91.00% and 90.94%, respectively, which are similar to the state-of-the-art results from a RNN [20], but also attention weights can be further used for visualisation making the model output interpretable which is the important and unique feature of the proposed model.

These learnings are beneficial for similar text classification research on other clinical tasks, such as cancer staging from pathology reports [10], diagnosis coding from medical records [11; 12], and prediction of mortality and unplanned readmissions [13]. It also sets a solid foundation for further improving performances on the "pain" models to scale the "pain" study to other hospitals and regions.

References

[1]   K.H. Todd, A review of current and emerging approaches to pain management in the emergency department, *J Pain Therapy* **6** (2017), 193-202.
[2]   Care Quality Commission., 2016 Emergency Department Survey: Statistical Release, in: Q.C. Commission, ed., National Health Service, London, 2016.
[3]   F. Karwowski-Soulié, S. Lessenot-Tcherny, A. Lamarche-Vadel, S. Bineau, C. Ginsburg, O. Meyniard, B. Mendoza, P. Fodella, G. Vidal-Trecan, and F. Brunet, Pain in an emergency department: an audit, *European Journal of Emergency Medicine* **13** (2006), 218-224.
[4]   W. Varndell, M. Fry, and D. Elliott, Quality and impact of nurse-initiated analgesia in the emergency department: A systematic review, *International emergency nursing* **40** (2018), 46-53.
[5]   C. Hatherley, N. Jennings, and R. Cross, Time to analgesia and pain score documentation best practice standards for the emergency department–a literature review, *Australasian Emergency Nursing Journal* **19** (2016), 26-36.
[6]   A.C. Williams and K.D. Craig, Updating the definition of pain, *J Pain* **157** (2016), 2420-2423.
[7]   L. Marquié, E. Raufaste, D. Lauque, C. Mariné, M. Ecoiffier, and P. Sorum, Pain rating by patients and physicians: evidence of systematic pain miscalibration, *Journal of Pain* **102** (2003), 289-296.
[8]   M. Duignan and V. Dunn, Congruence of pain assessment between nurses and emergency department patients: a replication, *International emergency nursing* **16** (2008), 23-28.
[9]   J.A. Hughes, N.J. Brown, J. Chiu, B. Allwood, and K. Chu, The relationship between time to analgesic administration and emergency department length of stay: A retrospective review, *Journal of Advanced Nursing* (2019).
[10] I. McCowan, D. Moore, and M.-J. Fry, Classification of cancer stage from free-text histology reports, in: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2006, pp. 5153-5156.
[11] B. Koopman, G. Zuccon, A. Nguyen, A. Bergheim, and N. Grayson, Automatic ICD-10 classification of cancers from free-text death certificates, *International journal of medical informatics* **84** (2015), 956-965.

[12] J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein, Explainable prediction of medical codes from clinical text, *Proceedings of the 2018 Conference of the North American Chapter* (2018).

[13] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, P.J. Liu, X. Liu, J. Marcus, and M. Sun, Scalable and accurate deep learning with electronic health records, *Digital Medicine* **1** (2018), 18.

[14] Z. Che, J.S. Sauver, H. Liu, and Y. Liu, Deep learning solutions for classifying patients on opioid use, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2017, p. 525.

[15] B. Scholkopf and A.J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2001.

[16] A. Liaw and M. Wiener, Classification and regression by Random Forest, *R News* **2** (2002), 18-22.

[17] J.L. Elman, Finding structure in time, *Cognitive science* **14** (1990), 179-211.

[18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86** (1998), 2278-2324.

[19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, *NIPS 2014 Workshop on Deep Learning* (2014).

[20] T. Vu, A. Nguyen, N. Brown, and J. Hughes, Identifying patients with pain in emergency departments using conventional machine learning and deep learning, in: *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, 2019, pp. 111-119.