

Natural language processing - pressure injury

Project report

Intensive Care NSW

October 2024

The information in this resource should not replace a clinician's professional judgement.

The Agency for Clinical Innovation (ACI) works with clinicians, consumers and system leaders to improve the health system in NSW. We design and implement new ways to deliver care for patients.

The ACI partners with the Ministry of Health, pillar organisations, local health districts, specialty networks, clinicians, patients and carers. We work together to achieve better value in healthcare by:

- improving patient and provider experience
- improving patient outcomes
- delivering efficient and sustainable healthcare services.

The ACI is part of the Clinical Innovation and Research Division and works closely with the Office for Health and Medical Research.

aci.health.nsw.gov.au

AGENCY FOR CLINICAL INNOVATION

1 Reserve Road St Leonards NSW 2065
 Locked Bag 2030, St Leonards NSW 1590
 Phone +61 2 9464 4666
 Email aci-info@health.nsw.gov.au | aci.health.nsw.gov.au

Further copies of this publication can be obtained from the Agency for Clinical Innovation website at aci.health.nsw.gov.au

Disclaimer: Content within this publication was accurate at the time of publication.

© State of New South Wales (Agency for Clinical Innovation) Year. Creative Commons Attribution-No Derivatives 4.0 licence. For current information go to aci.health.nsw.gov.au. The ACI logo is excluded from the Creative Commons licence and may only be used with express permission.

Title	Natural language processing- pressure injury		
Published	October 2024		
Next review	2029		
Produced by	Intensive Care NSW		
Preferred citation	NSW Agency for Clinical Innovation. Natural language processing- pressure injury. Sydney: ACI; 2024.		
CM ACI/D24/3057	SHPN (ACI) 241050	ISBN 978-1-76023-735-6	ACI_10915

Contents

- Background..... 4**
- Aim and objectives 5**
- Project overview 5**
- Project outcomes 6**
 - Outcome 1: Literature review 6
 - Outcome 2: Classification algorithm..... 7
 - Outcome 3: Prediction algorithm 11
 - Outcome 4: Algorithms for other hospital acquired complications..... 13
 - Outcome 5: Implementation plan 13
- Lessons learned and recommendations 14**
 - Challenges 14
 - Discussion 14
 - Recommendations 15
- Appendix 16**
 - Logistic regression model coefficients..... 16
 - Machine learning tuning parameters 17
- References 18**
- Glossary 19**
- Acknowledgements..... 21**

Background

Pressure injuries (PI) are an hospital-acquired complication which disproportionately affects patients in intensive care units (ICUs).¹ Development of a PI can lead to multiple adverse patient outcomes, including increased levels of pain, longer hospital length of stay and overall reduced quality of life. Given that many PIs are preventable, it is vital that accurate, real-time surveillance of PIs be performed at local and system-wide levels to identify and address any negative change in trends early. Despite a range of PI prevention strategies being implemented as standard care across Australian ICUs, PIs still occur. One recent study reported a point prevalence rate of 9.7% in Australian adult ICU patients.^{2, 3} By identifying and targeting these high-risk patients for additional preventative measures, the incidence of PIs could be reduced, thereby delivering benefits to the patients and the healthcare system.

Pressure injury surveillance systems and risk stratification tools currently exist, however both are associated with significant limitations, particularly when applied to the ICU context. Several different types of PI surveillance systems also exist internationally. Two common surveillance types are:

- incident reporting systems, which rely on clinicians entering a notification when they identify a PI
- administrative data surveillance systems, which typically rely on the clinical coding summary produced at the end of each inpatient episode of care.^{4, 5}

Multiple studies have reported the disadvantages associated with these two systems, including poor data quality and time delays required to generate the clinical coding summary.^{6, 7} Similarly, PI risk assessment tools have been repeatedly shown to perform poorly in the ICU population, with a tendency to classify all ICU patients as high risk.⁸

Natural language processing (NLP) is the branch of artificial intelligence that allows computers to interpret and analyse text-based data. It has previously been used to extract information from clinical notes for a wide range of conditions, including obesity, diabetes and mental health disorders. Although very few studies have been conducted in Australia.⁹⁻¹¹ Using NLP to extract information about PIs directly from the clinical progress notes could generate a more complete dataset in near real-time, thereby mitigating many of the problems inherent in current surveillance systems. Furthermore, the data from this system could then also be used to develop a PI risk assessment tool specific to the adult ICU population, allowing preventative strategies and a more targeted approach to the management of PIs.

This report details the findings from a project conducted by the Agency for Clinical Innovation (ACI) Intensive Care NSW (ICNSW) between February 2023 and July 2024 which investigated the feasibility of using NLP in the ICU clinical context.

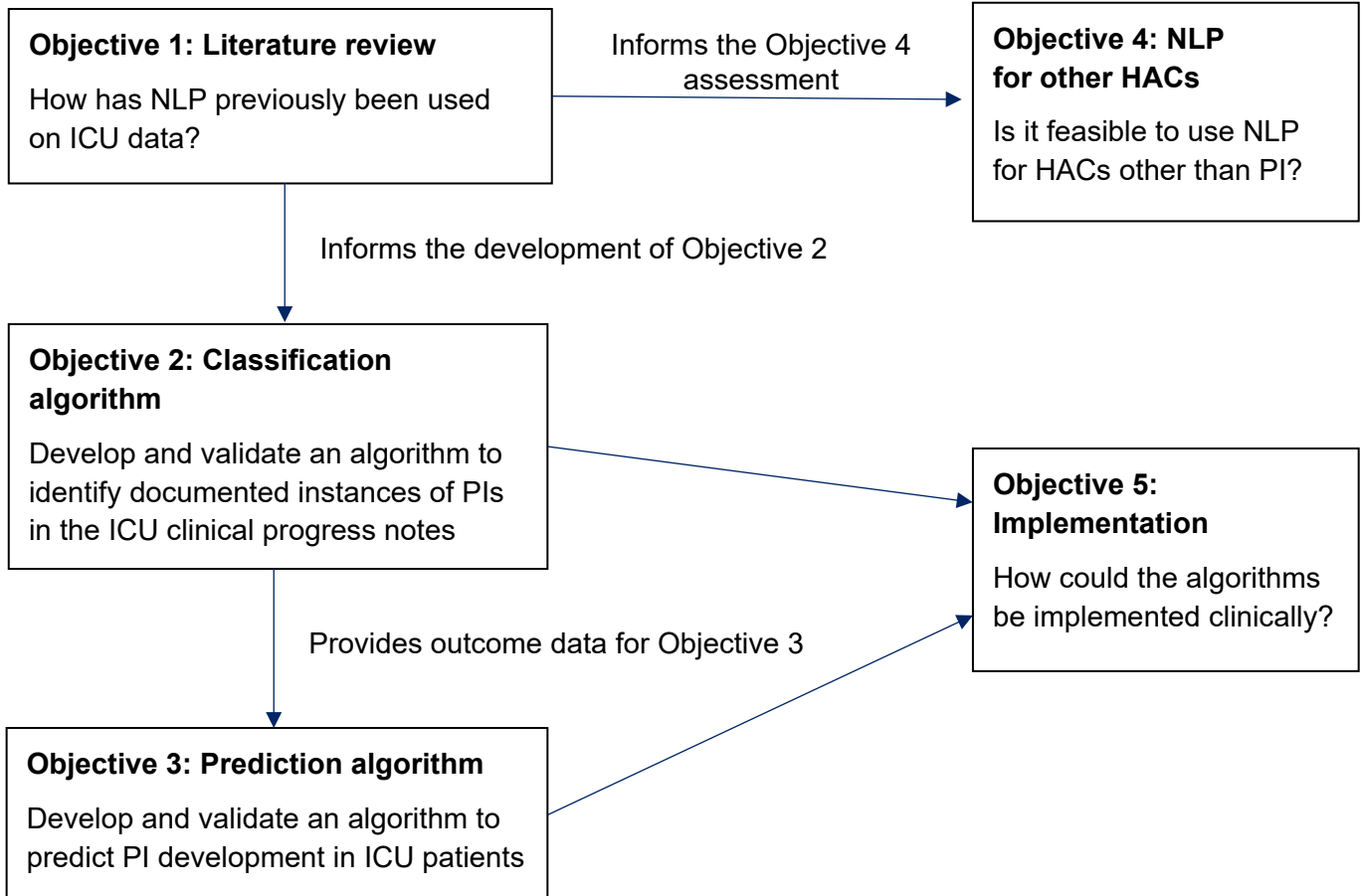
Aim and objectives

The overall aim of the project was to assess the feasibility of using routinely collected, unstructured data within the ICU clinical information system (CIS) to perform PI surveillance and risk stratification in the NSW adult ICU population. This aim was supported by a series of objectives to assess feasibility through an examination of previous research in this area and through the development of pilot algorithms using data from NSW:

1. Perform a literature review to identify studies which had previously used NLP on data generated in the ICU.
2. Develop and validate an algorithm to classify PIs documented in the ICU clinical progress notes.
3. Develop and validate an algorithm to predict PI development in adult ICU patients.
4. Scope the feasibility of using NLP techniques for other types of hospital acquired complications (HACs).
5. Develop a plan to implement the algorithms in a clinical setting.

Project overview

Figure 1: Project outcomes process



Project outcomes

This section of the report describes the outcomes of each of the project objectives and how these findings relate to the overall project aim.

Outcome 1: Literature review

The aim of the literature review was to provide an overview of how NLP has previously been used in the ICU clinical context and identify opportunities for how these techniques may be used in the future. The full review has been published in a **peer-reviewed journal**, with a summary of the review methods and findings outlined below.¹¹

Methods

A scoping review methodology was chosen to match the broad aim of the review and need for narrative synthesis given the likely heterogeneity of relevant articles.¹² To identify potentially relevant articles, a search strategy which combined the concepts 'natural language processing' and 'intensive care' was performed on the PubMed database on 9 March 2023. Title and abstract screening was performed by two members of the project team, as was subsequent full text review and data extraction. A narrative synthesis was performed using a framework based on the broad categories of study type identified in the review and each study was evaluated to determine the described algorithms' level of readiness for implementation into a clinical setting.^{13, 14}

Results

A total of 87 studies were included in the review. More than two-thirds of the included studies used the publicly available Medical Information Mart for Intensive Care dataset (67.8%, n=59) and only two studies used Australian data. The most common types of study were those which aimed to predict a clinical outcome (27.6%, n=24) or which aimed to develop an algorithm capable of extracting information about a particular clinical concept or condition from the progress notes (26.4%, n=23). The level of readiness for implementation of algorithms described in the identified studies was generally low, with the majority only reporting algorithm development and internal validation (90.8%, n=79). Only two studies had been prospectively tested in a clinical setting. Most studies were written for an audience with a computer science or health informatics background, with relatively few published in journals intended for intensive care clinicians.

Overall, the findings of the review highlighted that using NLP on ICU progress notes is technically feasible, supporting the use of NLP to develop the algorithms described later in this report. However, there has been little clinician involvement in this area and the clinical utility of these types of algorithms is unknown.

Outcome 2: Classification algorithm

The aim of Outcome 2 was to develop an algorithm capable of performing accurate, system-wide surveillance of PI development in the adult ICU population and to compare its performance against clinical coding data and an assessment by an expert clinician. All data for this part of the project was extracted from the Electronic Record for Intensive Care (eRIC) CIS via the NSW Health Enterprise Data Lake (EDL). A summary of how the classification algorithms were developed and validated is presented below.

Methods

Rules-based algorithm development

To develop the initial rules-based algorithm, text in the ‘Nursing – Integumentary Assessment’ progress note field was extracted from 574 patients known to have developed a PI at Site A (16,716 notes). The rules-based algorithm used two dictionaries of terms: true and false positives. The true positive dictionary contained terms that were likely indicative of the presence of a PI such as ‘pressure injury’ or ‘pressure area’ whereas the false positive dictionary contained negations, such as ‘nil pressure injury’ and other false positive phrases such as ‘pressure area care’. The algorithm then counted the number of times these terms occurred in each progress note. If the number of true positive terms was greater than the number of false positive terms, then the note was classified as referring to a PI. The results of the individual note classifications were then aggregated to determine each patient’s overall PI classification.

Rules-based algorithm overview

Classification rule: If ‘True positive count’ > ‘False positive count’ then classify as pressure injury.

True positive terms

- pressure injury
- pressure area
- PI
- PA

False positive terms

- nil pressure injury
- pressure area care
- no PI present
- PA catheter

Example

Note name	Text	True positive count	False positive count	Pressure injury present?
Note A	Pressure area care maintained. Stage 2 PI to sacrum noted.	2	1	Is 2 > 1? → Yes
Note B	Patient given full wash. Regular pressure area care attended.	1	1	Is 1 > 1? → No
Note C	PA catheter not working - take out.	1	1	Is 1 > 1? → No

Machine learning algorithm development

Machine learning models have also been developed to classify the progress note entries. Tree-based algorithms were used, with the decision tree algorithm yielding good compromise of classification accuracy as well as computation time. More complex algorithms, such as boosted tree, provided some improvement in accuracy, but at significantly longer computation time (e.g. training the decision tree algorithm took 40 minutes while training the boosted tree algorithm was aborted after 30 hours). Reported metrics in this report are from decision tree models.

The rules-based algorithm generated significantly more training data to be leveraged by the decision tree algorithm. Table 1 outlines the number of annotated notes and words in the vocabulary provided by manual and rules-based algorithm annotation.

Table 1: Amount of training data provided by manual and rules-based algorithm annotations

Annotation method	Number of annotated notes	Number of words in the vocabulary
Manual annotation	1,191	43,000
Rules-based algorithm	16,000	387,000

Algorithm validation

The performance of the various algorithms was assessed using a combination of manual chart review, incident management system data and clinical coding data. Relative performance of clinical coding data compared to the algorithms was also assessed where available. This was assessed by using an ICD-10-AM based algorithm to identify PIs in ICU patients and quantifying its performance against the same reference standard as the NLP algorithm. The ICD algorithm consisted of searching for any code containing 'L89' (pressure ulcer) in the clinical coding summary for all patients who also had an ICU location registered during their episode of care, as has been done previously.⁴ At one site, a comprehensive chart review was also performed by an expert clinician to generate an expert clinical assessment reference standard which enabled the quantification of errors written in the progress notes. Performance of the algorithm in correctly classifying individual progress notes was also evaluated through manual review of a sample of 'Nursing – Integumentary Assessment' notes from two sites (A and B). A summary of the various reference standards used to evaluate the algorithms' ability to determine whether a patient had a PI documented during their ICU admission is presented in Table 2.

Table 2: Description of the various reference standards used to assess algorithm performance

Reference standard	Description	Site
Manual chart review	At sites where no secondary data source was available, a random sample of patients was selected and their charts reviewed to assign the reference label.	Site B
Manual chart review + incident data	Classifications made by the algorithm were compared to those in the incident data. Where the two methods agreed, this was taken as the reference label and where they disagreed a manual chart review of the ICU progress notes was performed to assign the reference label.	Site A
Manual chart review + clinical coding data	Classifications made by the algorithm were compared to those in the clinical coding summary. Where the two methods agreed, this was taken as the reference label and where they disagreed a manual chart review of the ICU progress notes was performed to assign the reference label.	Sites C, D, E, F
Expert clinical assessment	As for the manual chart review + incident data method, except the manual reviews were conducted by a senior clinician who reviewed the whole patient record as well as local incident data to determine if the patient had developed a PI.	Site A

Results

Rules-based algorithm

The rules-based algorithm was run on all free-text fields for all patients who had their episode of care recorded in the ICU CIS at participating sites from 1 January 2017 to 30 June 2023 (Table 3). In total, this equated to 177,549,006 free-text fields across 54,111 episodes of care.

Table 3: Number of episodes of care and free text fields used to evaluate the rules-based algorithm

Site	Number of episodes of care	Number of free text field values
Site A	15,788	49,555,318
Site B	12,070	44,489,183
Site C	4,734	17,796,891
Site D	2,562	6,680,092
Site E	16,286	51,194,282
Site F	2,671	7,833,240
Total	54,111	177,549,006

The performance of the rules-based algorithm against the various reference standards is presented in Table 4 as is the relative performance of the ICD-10-AM code-based algorithm against the same reference standard.

Table 4: Performance metrics for the rules-based algorithm against a range of reference standards and relative performance of the clinical coding algorithm

Reference standard	Accuracy	F1 score	Sensitivity	Specificity
NLP rules-based algorithm				
Manual chart review (Site A)*	0.959	0.940	0.951	0.963
Manual chart review + incident data (Site A)†	0.968	0.848	0.961	0.968
Expert clinical assessment (Site A)†	0.958	0.791	0.956	0.958
Manual chart review (Site B)*	0.968	0.833	0.984	0.966
Manual chart review (Site B)†	0.773	0.714	0.966	0.693
Manual chart review + clinical coding data (Sites C and D)†	0.954	0.749	0.998	0.951
Manual chart review + clinical coding data (Sites E and F)†	0.947	0.743	0.961	0.945
ICD-10-AM algorithm				
Manual chart review + clinical coding data (Sites and D)†	0.948	0.322	0.210	0.993
Manual chart review + clinical coding data (Sites E and F)†	0.946	0.611	0.523	0.983

* Individual progress note level classification

† Patient level classification

Overall, these performance metrics (F1 score ranging from 0.940 to 0.714 across methods and sites) demonstrate feasibility of using NLP to identify documented instances of PIs in the ICU clinical progress notes. These metrics of the rules-based algorithm were superior to those of the clinical coding-based algorithm (F1 score 0.749 vs 0.322 at sites C and D and 0.743 vs 0.611 at sites E and F) and were also superior to those of clinical coding and incident management data reported in previous studies.^{4, 5} This suggests that using NLP to generate surveillance data in the clinical setting may result in more accurate data than existing systems.

Machine learning algorithm

The free text notes classified by the rules-based algorithm also served as the training data for machine learning algorithms. A decision tree model trained on manually annotated data (1191 notes) resulted in F1 score of 0.68 while the same type of model trained on rules-based classified training data (16,000 notes) resulted in 0.84 during internal validation (i.e. 80% of data used for training and 20% for validation).

Table 5: Performance metrics for the decision tree algorithm using different training data

Training data	Balanced accuracy	F1 score	Sensitivity	Specificity
Manual chart review	0.78	0.68	0.95	0.62
Rules-based algorithm annotation	0.92	0.84	0.94	0.90

The performance of the decision tree algorithm against the reference standards are also noted in Table 6.

Table 6: Performance metrics for the decision tree algorithm against a range of reference standards - Machine learning algorithm

Reference standard	Accuracy	F1 score	Sensitivity	Specificity
Manual chart review + incident data (Site A)‡	0.84	0.56	0.83	0.98
Expert clinical assessment (Site A)‡	0.53	0.69	0.01	1.00
Manual chart review (Site B)‡	0.73	0.68	0.61	1.00
Manual chart review + clinical coding data (Sites C and D)‡	0.78	0.43	0.76	0.98
Manual chart review + clinical coding data (Sites E and F)‡	0.78	0.45	0.76	0.95

‡ Patient level classification

The performance metrics of the machine learning algorithm (F1 score ranging from 0.43 to 0.68 across review method and sites) showed further tuning of the decision tree algorithm or usage of a more complex algorithm may be required for better accuracy.

Outcome 3: Prediction algorithm

The aim of this part of the project was to develop an algorithm capable of identifying adult ICU patients who are at high risk of PI development. Data for this project was extracted from the eRIC CIS via the NSW Health EDL and from the Australia and New Zealand Intensive Care Society (ANZICS) adult patient database (APD).

Methods

Outcome labels regarding whether a patient had developed a PI or not were obtained using the rules-based classification algorithm described in Outcome 2. To identify and exclude PIs likely acquired prior to ICU admission, the date and time of the first note identified as referring to a PI was extracted for each positive patient. Any patient who had this first note recorded within 24 hours of the ICU admission date and time was then excluded from the algorithm development dataset.

Candidate predictor variables were initially identified by consulting systematic reviews relating to risk factors for PI development in the ICU population as well as expert clinicians in this area. A detailed mapping of how these variables could be extracted from the ICU CIS was conducted, including

identification of which variables may require the use of NLP. However, as the project progressed, it became apparent that extraction of these variables via the ICU CIS would not be feasible to perform within the project's timeframe, due to difficulties associated with locating the data within the CIS. Therefore, the APD was identified as an alternative data source for the predictor variables.

Results

A range of different model types were tested, the performance of which is detailed in Table 7. Details of the coefficients of the logistic model and the tuning parameters (hyperparameters) of the various machine learning models are presented in the [Appendix](#).

Table 7: Performance of the predictive algorithms using data from the ANZICS APD

Model type	Accuracy	Balanced accuracy	Sensitivity	Specificity	Area under the receiver operating characteristic curve
Logistic regression [§]	0.93	0.510	0.02	1.00	0.750
Gradient boosted tree [§]	0.88	0.525	0.99	0.06	0.525
Decision tree [§]	0.86	0.550	0.95	0.14	0.550
Support vector machine [§]	0.88	0.500	1.00	0	0.500
K nearest neighbour [§]	0.88	0.510	0.99	0.01	0.510
Gradient boosted tree ^{**} – with resampling	0.70	0.700	0.69	0.71	0.700
Decision tree ^{2**} – with resampling	0.67	0.670	0.64	0.70	0.670

[§] No resampling used

^{**} Resampling used to obtain prevalence of 50%

With the highest balanced accuracy being 0.70, the performance of these algorithms is not sufficient to consider implementing them in the clinical setting. The poor performance is likely due to two factors. First, the dataset used to train and test the models was highly imbalanced. The low prevalence (11.8%) of PI in the training data resulted in the machine learning algorithms to be trained with more non-PI data than PI data. Using resampled training data with 50% PI prevalence improved the specificity and therefore balanced accuracy. However, the resultant models continued to demonstrate unsatisfactory results. The second factor relates to a limitation of the ANZICS APD, namely that the data contained within it is restricted to the first 24 hours of the patients' ICU admission. Incorporating more longitudinal data from the duration of the ICU admission, along with other variables not available in the APD, may improve the algorithms' performance.

Outcome 4: Algorithms for other hospital acquired complications

This part of the project's objective was to evaluate which other types of HAC may be feasible to develop a type of NLP algorithm for. Multiple studies identified in the scoping review described algorithms for a range of HACs, including catheter associated urinary tract infections, healthcare associated infections, falls, pneumonia, acute kidney injury, bleeding and adverse drug events. Another systemic review conducted specifically to identify studies which either used NLP to identify adverse events or extract information from incident reports identified an additional 35 studies.¹⁵ This vast amount of literature indicates that NLP is clearly feasible for use in identifying a range of HACs and could potentially be used as an alternative surveillance tool in the future. However, like the broader finding of the scoping review, studies in this area are predominantly retrospective and the clinical impact these algorithms would have once implemented remains unknown.

Outcome 5: Implementation plan

Previous sections of this report describe the feasibility and testing of the classification and prediction algorithms. Based on these findings, the project team concluded that the classification algorithm performed to a high standard, detecting 97% of all PIs which occurred in the cohort, whereas the prediction algorithm requires further development.

As previously outlined, there are several disadvantages inherent in the current systems used to monitor PIs in the ICU, namely low data quality, time delays associated with data availability and increased documentation burden for clinical staff.⁷ Implementing the classification algorithm as part of a new ICU PI monitoring system could potentially increase data quality and reduce clinician documentation burden. The algorithm would present a daily summary of documented PIs to a nominated clinical expert. The expert would then review the patient and confirm if the documentation was accurate. The expert would also provide feedback to staff about the accuracy of their documentation and provide a management plan for any identified PIs. The expert would also report the PI in the Incident Management System (IMS) as per local policy. This will improve the quality of IMS data and eliminate the need for clinical staff to report the PI in multiple systems.

To implement this type of reporting process, the next step would be to develop a system which can present the output of the algorithm to the clinical expert. This would involve programming the algorithm to run regularly on the progress note data stored in the NSW Health EDL and present its output via an application such as Microsoft Power BI. The application would need to identify which patients had been flagged as having had a PI documented in their notes to enable the expert to perform a clinical review. Development of the application interface would also require input from potential end-users to ensure the information being presented can be easily interpreted. Once the system has been developed and tested, implementation and evaluation in the clinical setting can be considered.

Lessons learned and recommendations

Overall, the NLP PI project successfully met its aim of assessing the feasibility of using routinely collected unstructured data within the ICU CIS to perform PI surveillance and risk stratification in the NSW adult ICU population. The pilot surveillance algorithms performed well and are ready to be tested in a clinical setting. All objectives were completed within the timeframe described in the project plan.

Challenges

- Technical challenges were presented by using a new cloud data platform when extracting the eRIC data from EDL. The project team had not used the EDL (Snowflake) before and were required to become familiar with the technical differences from the on-premise Microsoft SQL Server, such as the SQL syntax difference. Furthermore, while the team contained expertise relating to the eRIC CIS database structure, the clinical coding data was stored in the Cerner CIS. The team was unfamiliar with that database.
- Similarly, although a cloud-based analytics platform is available in conjunction with the EDL, it was not feasible for the team to learn how to fully leverage this platform within the project's timeframe. Instead, all algorithms were trained locally on the teams' laptops which resulted in a much slower training time and meant the performance of more complex algorithms, such as neural networks, was not able to be evaluated.
- On starting the data extraction, various data quality issues were identified, particularly when attempting to link the eRIC data to the APD dataset. This related to duplication of records when an individual changed their name or medical record number during their admission.

Difficulties encountered in obtaining data in this project are reflective of the broader, system-wide problem relating to data governance and access for quality improvement and research projects. While the EDL theoretically makes a large proportion of CIS data from NSW available for these purposes, the technical and data quality obstacles associated with obtaining data from the EDL mean its use is prohibitive for many project teams.

Discussion

While the upcoming intensive care data asset may largely address some of these issues, it would be helpful to have a mechanism to allow all current and potential users of the EDL to easily identify which other groups are using the EDL and for what purpose.

The findings of the literature review and the pilot algorithms described in this report support the feasibility of using NLP for a range of clinical tasks such as surveillance. However, as the literature review also highlighted, very limited work has been done to implement and evaluate the clinical impacts of NLP algorithms. Clinical artificial intelligence in general has very limited uptake in Australia and the lack of implementation studies has been cited as a key factor behind this.¹⁶ Testing the implementation of the surveillance algorithm would provide a valuable test case for how these types of projects could occur in future while simultaneously evaluating a potentially useful clinical tool.

Recommendations

- Develop a central registry of previous projects which have used the EDL, with the associated codes and queries used.
- Investigate whether it is feasible to implement the pilot surveillance algorithms described here rather than attempting to develop additional algorithms for topics beyond PIs.

Appendix

Logistic regression model coefficients

Variable	Beta	Beta (95% CI low)	Beta (95% CI high)	OR	OR (95% CI low)	OR (95% CI high)
(Intercept)	-8.769	-14.308	-3.395	0	0	0.034
ICU admission source						
Operating theatre and recovery (reference standard)	N/A	N/A	N/A	N/A	N/A	N/A
Emergency department	-1.980	-4.135	1.084	0.138	0.016	2.956
Transfer within same hospital (ward or ICU)	-1.724	-3.884	1.342	0.178	0.021	3.827
Transfer from other hospital (ward or ICU)	-1.585	-3.758	1.486	0.205	0.023	4.419
Home	-12.087					
Diabetes	-0.021	-0.259	0.212	0.979	0.772	1.236
Elective	-0.510	-0.857	-0.165	0.600	0.424	0.848
Decile of relative socioeconomic advantage or disadvantage	-0.066	-0.106	-0.025	0.936	0.899	0.975
Age	-0.002	-0.008	0.005	0.998	0.992	1.005
Apache3 score	0.007	0.002	0.012	1.007	1.002	1.012
Invasive ventilation in first 24hrs	1.434	1.200	1.672	4.195	3.320	5.323
Highest temperature in first 24hrs	0.238	0.111	0.365	1.269	1.117	1.441
Lowest albumin in first 24hrs	-0.050	-0.070	-0.029	0.951	0.932	0.971
Highest urea in first 24hrs	0.021	0.006	0.036	1.021	1.006	1.037
Lowest arterial oxygen pressure in first 24hrs	-0.003	-0.005	-0.001	0.997	0.995	0.999
Surgical admitting diagnosis	-2.086	-4.241	0.975	0.124	0.014	2.651
ICU admitting diagnosis						

Variable	Beta	Beta (95% CI low)	Beta (95% CI high)	OR	OR (95% CI low)	OR (95% CI high)
Cardiovascular (Reference Standard)	N/A	N/A	N/A	N/A	N/A	N/A
Respiratory	0.487	0.149	0.824	1.627	1.161	2.280
Gastrointestinal	-0.371	-0.761	0.009	0.690	0.467	1.009
Neurological	0.431	0.055	0.800	1.539	1.057	2.226
Sepsis	0.134	-0.288	0.548	1.143	0.750	1.730
Trauma	0.690	0.317	1.062	1.994	1.373	2.892
Other	0.059	-0.324	0.433	1.061	0.723	1.542
Pre-ICU hours	0	0	0.001	1.000	1.000	1.001

Machine learning tuning parameters

Algorithm	Hyperparameters	Notes
Gradient-boosted tree	Learning rate = 0.1 Max depth = 11 Number of estimators = 2000	Same values used for resampled training dataset
Decision tree	Max depth = 10	Same values used for resampled training dataset
Support vector machine	Default values	
K nearest neighbour	Number of neighbours = 10	

References

1. Coyer F, Miles S, Gosley S, et al. Pressure injury prevalence in intensive care versus non-intensive care patients: A state-wide comparison. *Aust Crit Care*. 2017;30(5):244-50. DOI: 10.1016/j.aucc.2016.12.003
2. Coyer F, Chaboyer W, Lin F, et al. Pressure injury prevalence in Australian intensive care units: A secondary analysis. *Aust Crit Care*. 2022;35(6):701-8. DOI: 10.1016/j.aucc.2021.10.009
3. Levido A, Fulbrook P, Barakat-Johnson M, et al. Pressure injury prevention practice in Australian intensive care units: A national cross-sectional survey. *Aust Crit Care*. 2023;36(2):186-94. DOI: 10.1016/j.aucc.2021.11.004
4. Team V, Tuck M, Reeves J, et al. Pressure injury data in Australian acute care settings: A comparison of three data sets. *Int Wound J*. 2020;17(3):578-86. DOI: 10.1111/iwj.13320
5. Ho C, Jiang J, Eastwood CA, et al. Validation of two case definitions to identify pressure ulcers using hospital administrative data. *BMJ open*. 2017;7(8):e016438. DOI: 10.1136/bmjopen-2017-016438
6. Nurmambetova E, Pan J, Zhang Z, et al. Developing an Inpatient Electronic Medical Record Phenotype for Hospital-Acquired Pressure Injuries: Case Study Using Natural Language Processing Models. *JMIR AI*. 2023;2:e41264. DOI: 10.2196/41264
7. Barakat-Johnson M, Lai M, Barnett C, et al. Hospital-acquired pressure injuries: Are they accurately reported? A prospective descriptive study in a large tertiary hospital in Australia. *J Tissue Viability*. 2018;27(4):203-10. DOI: 10.1016/j.jtv.2018.07.003
8. Zhang Y, Zhuang Y, Shen J, et al. Value of pressure injury assessment scales for patients in the intensive care unit: Systematic review and diagnostic test accuracy meta-analysis. *Intensive Crit Care Nurs*. 2021;64:103009. DOI: 10.1016/j.iccn.2020.103009
9. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *J Biomed Inform*. 2017;73:14-29. DOI: 10.1016/j.jbi.2017.07.012
10. Ford E, Carroll JA, Smith HE, et al. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;23(5):1007-15. DOI: 10.1093/jamia/ocv180
11. Pilowsky JK, Choi J-W, Saavedra A, et al. Natural language processing in the intensive care unit: A scoping review. *Crit Care Resusc*. 2024. DOI: 10.1016/j.ccrj.2024.06.008
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32. DOI: 10.1080/1364557032000119616
13. van de Sande D, van Genderen ME, Huiskens J, et al. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med*. 2021;47(7):750-60. DOI: 10.1007/s00134-021-06446-7
14. Pons E, Braun LMM, Hunink MGM, et al. Natural Language Processing in Radiology: A Systematic Review. *Radiology*. 2016;279(2):329-43. DOI: 10.1148/radiol.16142770
15. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. *Int J Med Inform*. 2019;132:103971. DOI: 10.1016/j.ijmedinf.2019.103971
16. van der Vegt A, Campbell V, Zuccon G. Why clinical artificial intelligence is (almost) non-existent in Australian hospitals and how to fix it. *Med J Aust*. 2023;220(4):172-5. DOI: 10.5694/mja2.52195

Glossary

Term, acronym or abbreviation	Definition
ANZICS	Australian New Zealand Intensive Care Society
APD	Adult Patient Database
Area under the receiver operating characteristic curve	Area under the receiver operating characteristic curve. Describes the probability of a model or algorithm correctly distinguishing between a case and a non-case correctly, ranging from 0.5 to 1, with 1 indicating perfect performance (100% probability that the classification will be correct).
CI	Confidence interval
CIS	Clinical information system
EDL	Enterprise Data Lake
eRIC	Electronic record for intensive care
F1 Score	Calculated as the harmonic mean of precision and recall. Provides an overall metric of the performance of a model or algorithm, ranging from 0 to 1, with 1 indicating perfect performance.
HAC	Hospital acquired complication
ICD	International Statistical Classification of Diseases and Related Health Problems
ICD-10-AM	International Statistical Classification of Diseases and Health Related Problems, 10th Revision, Australian Modification
ICU	Intensive care unit
IMS	Incident management system
Internal validation	Internal validation tests how well a model or algorithm performs on the data from which it was derived. This is commonly done by splitting the available data into a training and testing set.
K	The k-nearest neighbours algorithm (k-NN) is a non-parametric supervised learning method
Machine learning algorithm	Algorithms which learn from patterns in a dataset to determine their output. Examples include decision trees, random forests, support-vector machines and artificial neural networks.
NLP	Natural language processing
NSW	New South Wales
OR	Odds ratio
PI	Pressure injury

Term, acronym or abbreviation	Definition
Rule-based algorithm	Algorithms which depend on a series of pre-specified 'if then' rules to determine their output. Commonly used in combination with regular expressions.
SQL	Structured query language
Unstructured data	Data without a fixed format, such as free text, audio or image data.

Acknowledgements

We would like to recognise and thank the following organisations and individuals who were involved in the development of this resource.

- Sarah Jones- Clinical Nurse Consultant ICU – St George Hospital
- Warren Eather- Data / Information and Communication Technology Manager ICU, St George Hospital
- South Eastern Sydney Local Health District
- Northern Sydney Local Health District
- Central Coast Local Health District