



Universiteit
Leiden
The Netherlands

Natural language processing for automated detection of incidental durotomy

Karhade, A.V.; Bongers, M.E.R.; Groot, O.Q.; Kazarian, E.R.; Cha, T.D.; Fogel, H.A.; ... ; Schwab, J.H.

Citation

Karhade, A. V., Bongers, M. E. R., Groot, O. Q., Kazarian, E. R., Cha, T. D., Fogel, H. A., ... Schwab, J. H. (2020). Natural language processing for automated detection of incidental durotomy. *The Spine Journal*, 20(5), 695-700. doi:10.1016/j.spinee.2019.12.006

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/4302496>

Note: To cite this publication please use the final published version (if applicable).

Clinical Study

Natural language processing for automated detection
of incidental durotomy

Aditya V. Karhade, BE^{a,b}, Michiel E.R. Bongers, MD^a,
Olivier Q. Groot, MD^a, Erick R. Kazarian, MD^{a,b},
Thomas D. Cha, MD, MBA^a, Harold A. Fogel, MD^a,
Stuart H. Hershman, MD^a, Daniel G. Tobert, MD^a,
Andrew J. Schoenfeld, MS, MSc^b, Christopher M. Bono, MD^a,
James D. Kang, MD^b, Mitchel B. Harris, MD^a, Joseph H. Schwab, MD, MS^{a,*}

^a Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, 55 Fruit St, Boston, MA 02114, USA

^b Department of Orthopedic Surgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115, USA

Received 16 October 2019; revised 18 November 2019; accepted 16 December 2019

Abstract

BACKGROUND: Incidental durotomy is a common intraoperative complication during spine surgery with potential implications for postoperative recovery, patient-reported outcomes, length of stay, and costs. To our knowledge, there are no processes available for automated surveillance of incidental durotomy.

PURPOSE: The purpose of this study was to develop natural language processing (NLP) algorithms for automated detection of incidental durotomies in free-text operative notes of patients undergoing lumbar spine surgery.

PATIENT SAMPLE: Adult patients 18 years or older undergoing lumbar spine surgery between January 1, 2000 and June 31, 2018 at two academic and three community medical centers.

OUTCOME MEASURES: The primary outcome was defined as intraoperative durotomy recorded in free-text operative notes.

METHODS: An 80:20 stratified split was undertaken to create training and testing populations. An extreme gradient-boosting NLP algorithm was developed to detect incidental durotomy. Discrimination was assessed via area under receiver-operating curve (AUC-ROC), precision-recall curve, and Brier score. Performance of this algorithm was compared with current procedural terminology (CPT) and international classification of diseases (ICD) codes for durotomy.

RESULTS: Overall, 1,000 patients were included in the study and 93 (9.3%) had a recorded incidental durotomy in the free-text operative report. In the independent testing set (n=200) not used for model development, the NLP algorithm achieved AUC-ROC of 0.99 for detection of durotomy. In comparison, the CPT/ICD codes had AUC-ROC of 0.64. In the testing set, the NLP algorithm detected 16 of 18 patients with incidental durotomy (sensitivity 0.89) whereas the CPT and ICD codes detected 5 of 18 (sensitivity 0.28). At a threshold of 0.05, the NLP algorithm had specificity of 0.99, positive predictive value of 0.89, and negative predictive value of 0.99.

FDA device/drug status: Not applicable.

Author disclosures: **AVK:** Nothing to disclose. **MERB:** Nothing to disclose. **OQG:** Nothing to disclose. **ERK:** Nothing to disclose. **TDC:** Consulting: GE Healthcare, NuVasive, K2M, Bio2. **HAF:** Nothing to disclose. **SHH:** Nothing to disclose. **DGT:** Nothing to disclose. **AJS:** Wolters Kluwer, Springer; Scientific Advisory Board/Other Office: JBJS; Research Support (Investigator Salary, Staff/Materials): CMS-OMH; Grants: DoD, OREF, NIH-NIAMS. **CMB:** Royalties: Wolters Kluwer; Consulting: JAAOS, AAOS, United Healthcare, NASS; Scientific Advisory Board/Other Office:

The Spine Journal/NASS. **JDK:** Nothing to disclose. **MBH:** Nothing to disclose. **JHS:** Consulting: Stryker.

Ethics Statement: This study was approved by our institutional review board.

*Corresponding author. Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, 55 Fruit St, Boston, MA 02114, USA. Tel.: (617) 543-5227; fax: (617) 726-7587.

E-mail address: jhschwab@mgh.harvard.edu (J.H. Schwab).

CONCLUSIONS: Internal validation of the NLP algorithm developed in this study indicates promising results for future NLP applications in spine surgery. Pending external validation, the NLP algorithm developed in this study may be used by entities including national spine registries or hospital quality and safety departments to automate tracking of incidental durotomies. © 2019 Elsevier Inc. All rights reserved.

Keywords: Artificial intelligence; Diagnosis; Dural tear; Durotomy; Machine learning; Natural language processing; Prediction; Spine

Introduction

Incidental durotomy is a common occurrence in spine surgery with implications for postoperative recovery, patient-reported outcomes, length of stay, and costs [1–4]. The transition to value-based health care in spine surgery has increased pressure on provider systems to standardize the measurement of postoperative outcomes, reduce variation, and improve reporting of quality and safety metrics [5]. Studies of national administrative databases have used procedural and diagnosis codes to identify incidental durotomy [1,6,7]. However, to our knowledge, there are no algorithms available for automated surveillance of incidental durotomy.

The purpose of this study was to develop natural language processing (NLP) algorithms for automated detection of incidental durotomy in free-text operative notes of patients undergoing lumbar spine surgery. A secondary purpose of this study was to compare the performance of the NLP algorithm to administrative Current Procedural Terminology (CPT) and International Classification of Diseases (ICD-9 and ICD-10) codes for detection of incidental durotomy.

Materials and methods

Guidelines

The following guidelines were followed: transparent-reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [8].

Data source

Institutional review board approval was granted for retrospective review of electronic health records (EHR). Patients from two academic medical centers and three community medical centers affiliated with one health-care entity were included in this analysis. The inclusion criteria for this study were: (1) age 18 years or older, (2) primary procedure of posterior lumbar decompression and/or fusion for lumbar disk herniation, spondylolisthesis, or spinal stenosis, (3) date of procedure between January 1, 2000 and June 31, 2018, and (4) free-text operative note from the first index lumbar procedure in our institution's electronic health records available for review. Patients who underwent procedures for trauma, malignancy, infection, pseudarthrosis, and deformity were excluded. Free-text operative notes from 1,000 patients meeting these criteria in our institutional database were reviewed.

Outcome

The primary outcome was intraoperative unintentional (incidental) durotomy recorded in free-text operative notes. Operative notes were reviewed by a researcher (AVK) who was not involved in the care of these patients and all reports flagged for incidental durotomy were confirmed with senior authors.

Data analysis

A stratified 80:20 split of the available population was done to create training and testing cohorts. Raw free-text operative notes were processed to obtain documents suitable for model development [9,10]. Raw free-text reports were scrubbed of redundant whitespace, line breaks, and stop words such as “to,” “the,” “and.” Words in these free-text reports were then stemmed, for example “leaking” and “leaked” were converted to “leak.” [Supplementary Table 1](#) shows an example excerpt of a raw free-text operative note and the final processed version used as input to model development. The bag-of-words data representation method was applied to the processed free-text reports to identify features (words) for model development. Bag-of-words technique can be conceptualized as the creation of a matrix with rows defined by the operative reports and columns defined by variables that indicate the frequency or presence of features (word/tokens) in the document. The term frequency-inverse document frequency technique was used to regularize these results by balancing the prevalence of rare and very common features.

An extreme gradient boosting (XGBoost) NLP machine learning algorithm was developed to detect incidental durotomy [11–13]. The NLP algorithm was developed on the training set (n=800) and assessed on the testing set (n=200) by the following metrics: discrimination (receiver operating curve [ROC], area under the receiver operating curve [AUC-ROC], precision-recall curve [PRC], area under the precision-recall curve [AUC-PRC], sensitivity [recall], specificity, positive-predictive value [PPV, also known as precision], negative-predictive value, F1-score, positive likelihood ratio [LR+], negative likelihood ratio [LR–]), calibration (calibration intercept, calibration slope), and overall performance (Brier score) [14–16]. The ROC is a plot of sensitivity and (1-specificity) versus the PRC is a plot of precision (PPV) and recall (sensitivity). The PRC is particularly helpful for examining

algorithm performance on imbalanced outcomes (outcomes that occur in a minority of the patients studied) [16,17]. AUC-ROC and AUC-PRC both range from 0 to a maximum value of 1. An algorithm with AUC-ROC=0.5 is no better than chance. F1-score is defined as the harmonic mean or the weighted average of precision (PPV) and recall (sensitivity). F1-scores range from 0 to a maximum of 1. LR+ is defined as sensitivity divided by (1-specificity) and LR– is defined as (1-sensitivity) divided by specificity. Useful tests have LR+ above 1 and LR– less than 1. Multiple probability thresholds were used to examine the performance of the algorithm. Brier score is defined as the mean squared difference between predictions and observed outcomes. Perfect models would have a Brier score of 0. The null model Brier score (score for an algorithm generating a predicted probability equal to the population prevalence of the outcome for every patient) was calculated to benchmark the algorithm Brier score.

Second, CPT and ICD codes [CPT=63707, 63709, 63710, ICD-9=349.3, ICD-10=G96.11, G97.4] were used to determine the performance of these administrative codes for detection of incidental durotomy. The performance of these codes was assessed by the same metrics as the NLP algorithm. Finally, explanations were generated for predictions generated by the NLP algorithm to allow clinicians to interpret the model findings. Global explanations were provided averaging across all patients to demonstrate the most important factors used for detection. Local explanations were implemented at the level of individual patient operative notes (local explanations) and used to identify which features (words or phrases) the NLP algorithm used to detect incidental durotomy [18]. The capacity for enhanced visualization was undertaken by generation of automated highlighting of features used for detection in free-text operative notes. Anaconda Distribution (Anaconda, Inc., Austin, TX, USA), R version (The R Foundation, Vienna, Austria), RStudio (RStudio, Boston, MA,

USA), and Python (Python Software Foundation, Wilmington, DE, USA) were used for data analysis.

Results

Overall, 1,000 patients were included in this study. The patients had a median age of 60 (interquartile range: 46–71) years and 459 (45.9%) were female. The rate of incidental durotomy in this cohort was 9.3% (n=93).

In the independent testing set (n=200), the NLP algorithm achieved AUC-ROC of 0.99 (Fig. 1, Left), AUC-PRC of 0.94 (Fig. 1, Right), Brier score of 0.01, calibration intercept of 1.39, and calibration slope of 1.17 for detection of incidental durotomy (Table 1). In comparison, the null model Brier score was 0.08. The AUC-ROC of the combined CPT and ICD codes was 0.64, AUC-PRC was 0.39, and Brier score was 0.07. In the testing set, the CPT and ICD codes only detected 5 of the 18 (sensitivity=0.28) incidental durotomies. However, the CPT and ICD codes had perfect PPV and specificity and generated no false positives (Table 2).

At a threshold of 0.05, the NLP algorithm had an F1-score of 0.89, specificity of 0.99, PPV of 0.89, and NPV of 0.99 in the independent testing set. The algorithm correctly identified 16 (true positives) of the 18 patients (sensitivity=0.89) who had incidental durotomies in the testing cohort and yielded two false positives. At a threshold of 0.50, algorithm correctly identified 15 (true positives) of the 18 patients (sensitivity=0.83) who had incidental durotomies in the testing cohort and had yielded no false positives (specificity=1, PPV=1). Decision curve analysis of the NLP showed that the algorithm yielded greater standardized net benefit at all thresholds compared with default strategies of changes made for no patients, changes for all patients, or changes based on the CPT and ICD codes (Fig. 2).

Global explanations for the NLP algorithm showed that the algorithm used clinically meaningful features (words)

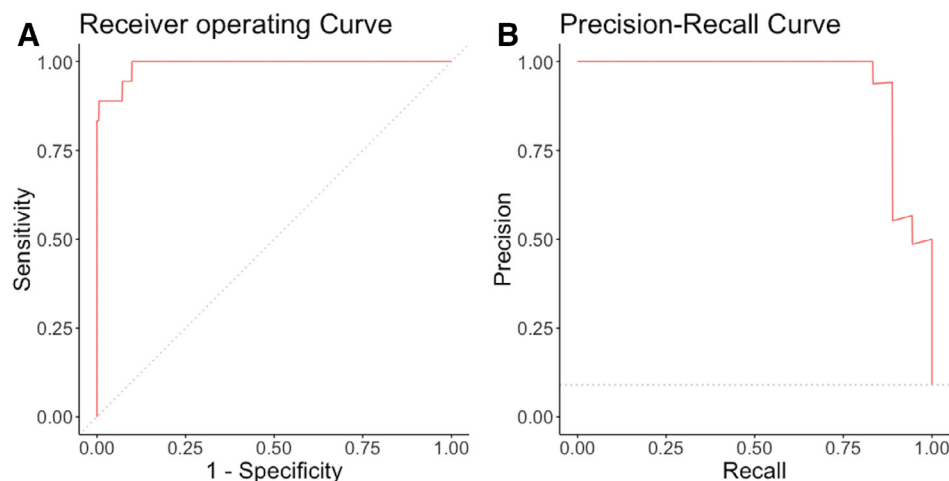


Fig. 1. (Left) Receiver operating curve and (Right) Precision-Recall curves for NLP algorithm in the independent testing set, n=200. NLP, natural language processing.

Table 1

Overall performance of NLP algorithm (95% confidence interval) relative to procedural and diagnosis codes in the independent testing set, n=200

	NLP algorithm	CPT and ICD codes
AUC-ROC	0.99 (0.96, 1.00)	0.64 (0.54, 0.75)
AUC-PRC	0.94 (0.75, 0.99)	0.39 (0.19, 0.60)
Brier	0.01 (0.005, 0.04)	0.07 (0.04, 0.10)
Calibration intercept	1.39 (0.26, 2.53)	-
Calibration slope	1.17 (0.50, 1.84)	-
null model Brier score=0.08		

AUC-PRC, area under the precision-recall curve; AUC-ROC, area under the receiver operating curve; CPT, current procedural terminology; ICD, international classification of disease; NLP, natural language processing.

for detection of incidental durotomy (“repair,” “leak,” “csf,” “tisseel,” “small_dural,” “leakag,” “prolen,” “dura”) (Fig. 3). Similarly, an example of the individual patient-level free-text operative report explanations of the NLP algorithm showed that “repair,” “small,” and “tisseel” in the operative note led the algorithm to detect the occurrence of incidental durotomy (Fig. 4).

Discussion

In this study, nearly one in 10 patients undergoing posterior lumbar decompression and/or fusion was found to have an incidental durotomy. On internal validation, the NLP algorithm developed in this study showed good performance for automated detection of incidental durotomy. Administrative codes were shown to have high positive predictive value but low sensitivity for automated surveillance of incidental durotomy.

Natural language processing has been previously applied to the automated extraction of structured data from unstructured free-text reports in radiology, pathology, and oncology, among others [19–21]. Within spine, Tan et al. previously studied machine learning-based and rules-based NLP for automated detection of degenerative lumbar spine findings (eg, deformity, fracture, disc disorders, and stenosis) in x-ray and magnetic resonance imaging radiology reports [22]. In separate work, Huhdanpaa et al. studied rules-based NLP for automated surveillance of Type 1

Table 2

Performance of NLP algorithm (95% confidence interval) at various thresholds relative to procedural and diagnosis codes in the independent testing set, n=200

	CPT and ICD codes	NLP algorithm		
		Threshold=0.50	Threshold=0.10	Threshold=0.05
Sensitivity (recall)	0.28 (0.10, 0.53)	0.83 (0.59, 0.96)	0.83 (0.59, 0.96)	0.89 (0.65, 0.99)
Specificity	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)	0.99 (0.97, 1.00)	0.99 (0.96, 1.00)
Negative predictive value	0.93 (0.89, 0.96)	0.98 (0.95, 1.00)	0.98 (0.95, 1.00)	0.99 (0.96, 1.00)
Positive predictive value (precision)	1.00 (0.48, 1.00)	1.00 (0.78, 1.00)	0.94 (0.70, 1.00)	0.89 (0.65, 0.99)
F1-score	0.43 (0.16, 0.70))	0.91 (0.67, 0.98)	0.88 (0.64, 0.97)	0.89 (0.65, 0.99)
LR (+)	-	-	150.8 (21.3, 1082.6)	80.9 (20.2, 324.1)
LR (-)	0.72 (0.54, 0.96)	0.17 (0.06, 0.47)	0.17 (0.06, 0.47)	0.11 (0.03, 0.42)

CPT, current procedural terminology; ICD, international classification of disease; LR-, negative likelihood ratio; LR+, positive likelihood ratio; NLP, natural language processing.

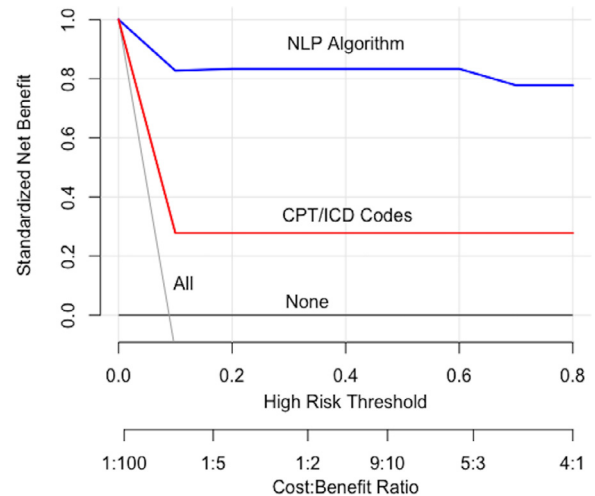


Fig. 2. Decision curve analysis with standardized net benefit of the NLP algorithm relative to default strategies of changes for no patients, or for all patients or changes based on the CPT/ICD codes. CPT, current procedural terminology; ICD, international classification of disease; NLP, natural language processing.

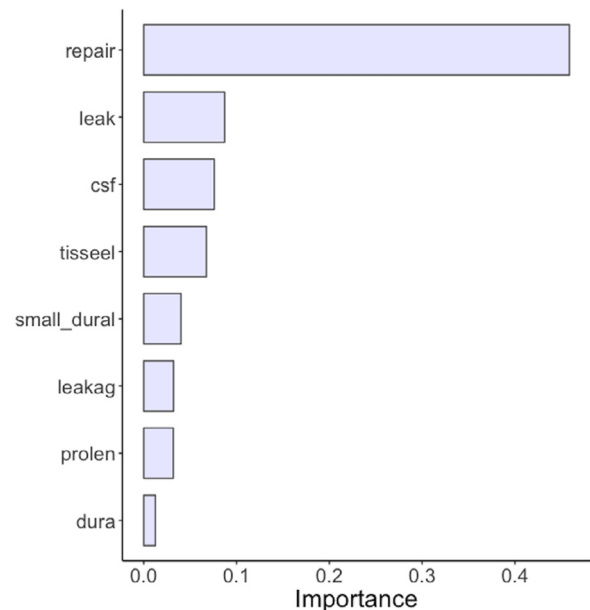


Fig. 3. Global explanation of features used by the NLP algorithm for detection of incidental durotomy. NLP, natural language processing.



Fig. 4. Example of individual patient-level explanation for incidental durotomy detection in free-text operative note generated by the NLP algorithm. NLP, natural language processing.

Modic changes in magnetic resonance imaging reports [23]. In both studies, surveillance of factors with relatively low prevalence (10% or less) yielded lower sensitivity than specificity, although multiple probability thresholds were not examined. With lower probability thresholds, the sensitivity of algorithms for the outcome increases while the specificity decreases, as demonstrated by the various thresholds examined in this analysis. Wyles et al. used rules-based NLP to identify operative approach, fixation technique, and bearing surface in free-text operative notes of patients undergoing total hip arthroplasty; however, standardized metrics for classification such as AUC-ROC were not reported [24]. Free-text operative notes have also been used for automated surveillance of intraoperative events such as periprosthetic femur fractures in total hip arthroplasty and posterior capsular rupture in cataract surgery [25,26].

The prevalence of incidental durotomy in this study was similar to previous studies of posterior lumbar decompression [27–31]. Studies that used procedural and diagnosis codes to identify incidental durotomy found a lower rate of incidental durotomy, consistent with the results presented here [3,6,7]. Desai et al. previously followed patients undergoing decompression and/or fusion for spinal stenosis from the Spine Patient Outcomes Research Trial and showed that there was no difference in mortality, reoperation, or

patient-reported outcomes up to 4-years after surgery who had an incidental durotomy [29]. However, patients with incidental durotomy had increased operative duration and inpatient stay. Similarly, Schroeder et al. showed that patients with incidental durotomy had 18% increase in average hospital charge despite matching to controls by age, Charlson comorbidity index, region of spine, type of fusion, number of fused levels, number of levels decompressed, and use of biologics [4]. Accurate depiction of rates of incidental durotomy thus has implications for value-based care delivery and alternative payment models.

There are a number of limitations that must be acknowledged to appropriately interpret the results of this study. This was a retrospective study from one health-care system beholden to the limitations inherent to such a research design. The surgeons affiliated with this health-care entity likely share practices that influence the specific terminology used to document durotomy. Prospective, temporal, and external validation of this algorithm remains to be performed. Additionally, there may be patients who had an incidental durotomy that was unrecognized or unrecorded in the operative note for the index procedure. Furthermore, repeated reviews of operative notes by multiple researchers or fellowship-trained spine surgeons could be preferable, although somewhat impractical, and this represents an added limitation of the current work.

Despite these limitations, this study has several important implications for quality and safety tracking and outcomes research in spine surgery. If found to be externally valid, the algorithm developed in this study could be used by professional societies for automated detection of incidental durotomies in national registries. Furthermore, quality and safety departments within specific health-care systems could use this algorithm to automatically track incidental durotomies over time and create metrics and dashboards for spine surgeons to monitor this complication. Finally, the efficiency of clinical outcomes research can be improved by using this algorithm to automate the identification of durotomy or similar complications in spine surgical cohorts. These aspirational goals rely on further refinement, characterization, and improvement of currently available NLP tools in spine surgery.

Conclusions

Internal validation of the NLP algorithm developed in this study indicates promising results for future NLP applications in spine surgery. Pending external validation, the NLP algorithm developed in this study may be used by entities including national spine registries or hospital quality and safety departments to automate tracking of incidental durotomies.

Acknowledgment

The authors report no funding disclosures for this study

Supplementary materials

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.spinee.2019.12.006>.

References

- [1] Nandyala SV, Elboghady IM, Marquez-Lara A, Noureldin MN, Sankaranarayanan S, Singh K. Cost analysis of incidental durotomy in spine surgery. *Spine* 2014;39(17):E1042–51.
- [2] Puvanesarajah V, Hassanzadeh H. The true cost of a dural tear: medical and economic ramifications of incidental durotomy during lumbar discectomy in elderly medicare beneficiaries. *Spine* 2017;42(10):770–6.
- [3] Buck JS, Yoon ST. The incidence of durotomy and its clinical and economic impact in primary, short-segment lumbar fusion: an analysis of 17,232 cases. *Spine* 2015;40(18):1444–50.
- [4] Schroeder GD, Kepler CK, Alijanipour P, Kurd MF, Hollern D, Kumar P, et al. The economic implications of an incidental durotomy. *Spine* 2016;41(19):1548–53.
- [5] Porter ME. What is value in health care? *New Eng J Med* 2010;363(26):2477–81.
- [6] Joseph JR, Smith BW, Park P. Variability in standard outcomes of posterior lumbar fusion determined by national databases. *World Neurosurg* 2017;97:236–40.
- [7] Passias PG, Horn SR, Jalai CM, Poorman G, Bono OJ, Ramchandran S, et al. Comparative analysis of perioperative complications between a multicenter prospective cervical deformity database and the Nationwide Inpatient Sample database. *Spine J* 2017;17(11):1633–40.
- [8] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13(1):1.
- [9] Loper E, Bird S. NLTK: the natural language toolkit. *arXiv preprint cs/0205028*. 2002.
- [10] Selivanov D, Wang Q. *text2vec: Modern Text Mining Framework for R*. Computer Software Manual] (R Package Version 04 0). Available at: <https://CRAN.R-project.org/package=text2vec>. 2016.
- [11] Xgboost: a scalable tree boosting system. In: Chen T, Guestrin C, editors. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM; 2016.
- [12] Torlay L, Perrone-Bertolotti M, Thomas E, Baci M. Machine learning—XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform* 2017;4(3):159.
- [13] Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C, et al. Prediction of the 1-year risk of incident lung cancer: prospective study using electronic health records from the state of maine. *J Med Internet Res* 2019;21(5):e13260.
- [14] Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54(1):17–23.
- [15] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35(29):1925–31.
- [16] Saito T, Rehmsmeier M. Precrec: fast and accurate precision–recall and ROC curve calculations in R. *Bioinformatics* 2017;33(1):145–7.
- [17] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One* 2015;10(3):e0118432.
- [18] Why should i trust you?: explaining the predictions of any classifier. In: Ribeiro MT, Singh S, Guestrin C, editors. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM; 2016.
- [19] Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. *Radiology* 2016;279(2):329–43.
- [20] Yim WW, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. *JAMA Oncol* 2016;2(6):797–804.
- [21] Burger G, Abu-Hanna A, de Keizer N, Cornet R. Natural language processing in pathology: a scoping review. *J Clin Pathol* 2016;11(11):949–55.
- [22] Tan WK, Hassanpour S, Heagerty PJ, Rundell SD, Suri P, Huhdanpaa HT, et al. Comparison of natural language processing rules-based and machine-learning systems to identify lumbar spine imaging findings related to low back pain. *Acad Radiol* 2018;25(11):1422–32.
- [23] Huhdanpaa HT, Tan WK, Rundell SD, Suri P, Chokshi FH, Comstock BA, et al. Using Natural Language Processing of Free-Text Radiology Reports to Identify Type 1 Modic Endplate Changes. *J Digital Imaging* 2018;31(1):84–90.
- [24] Wyles CC, Tibbo ME, Fu S, Wang Y, Sohn S, Kremers WK, et al. Use of natural language processing algorithms to identify common data elements in operative notes for total hip arthroplasty. *J Bone Joint Surg Am Vol* 2019;101(21):1931–8.
- [25] Liu L, Shorstein NH, Amsden LB, Herrinton LJ. Natural language processing to ascertain two key variables from operative reports in ophthalmology. *Pharmacoepidemiol Drug Saf* 2017;26(4):378–85.
- [26] Tibbo ME, Wyles CC, Fu S, Sohn S, Lewallen DG, Berry DJ, et al. Use of natural language processing tools to identify and classify periprosthetic femur fractures. *J Arthroplasty* 2019;34(10):2216–9.
- [27] Ishikura H, Ogihara S, Oka H, Maruyama T, Inanami H, Miyoshi K, et al. Risk factors for incidental durotomy during posterior open spine surgery for degenerative diseases in adults: a multicenter observational study. *PloS One* 2017;12(11):e0188038.
- [28] Ghobrial GM, Theofanis T, Darden BV, Arnold P, Fehlings MG, Harrop JS. Unintended durotomy in lumbar degenerative spinal surgery: a 10-year systematic review of the literature. *Neurosurg Focus* 2015;39(4):E8.
- [29] Desai A, Ball PA, Bekelis K, Lurie J, Mirza SK, Tosteson TD, et al. SPORT: does incidental durotomy affect long-term outcomes in cases of spinal stenosis? *Neurosurgery* 2015;76(Suppl 1):S57–63. discussion S.
- [30] Herren C, Sobottke R, Mannion AF, Zweig T, Munting E, Otten P, et al. Incidental durotomy in decompression for lumbar spinal stenosis: incidence, risk factors and effect on outcomes in the Spine Tango registry. *Eur Spine J* 2017;26(10):2483–95.
- [31] Ulrich NH, Burgstaller JM, Brunner F, Porchet F, Farshad M, Pichierri G, et al. The impact of incidental durotomy on the outcome of decompression surgery in degenerative lumbar spinal canal stenosis: analysis of the Lumbar Spinal Outcome Study (LSOS) data—a Swiss prospective multi-center cohort study. *BMC Musculoskelet Disord* 2016;17:170.