Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240822

Performance of a NLP Tool for Text Classification from Orthopaedic Operative Reports, Using Data from the Large Network of Clinical Data Warehouses of the West of France: The HACRO-HUGORTHO Project

Marie ANSOBORLO[a,b,1], Jeremy CARDOSO[a], Julien HERBERT[a], Christine SALPETRIER[a], Guillaume BOUZILLE[c], Marc CUGGIA[c], Philippe ROSSET[b,d], Louis-Romée LE NAIL[b,d] and Leslie GRAMMATICO-GUILLON[a,b]

[a] *Centre de données cliniques– Pôle santé publique prévention - CHRU de Tours, F-37000, Tours, France*

[b] *Faculté de Médecine, Université François-Rabelais de Tours, F-37000, France*

[c] *Univ Rennes, CHU Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France*

[d] *Service de Chirurgie Orthopédique et Traumatologique 2, Hôpital Trousseau, CHRU de Tours, F-37044, Tours, France*

ORCiD ID: Marie Ansoborlo https://orcid.org/0000-0002-6543-2301

Abstract. Electronic health data concerning implantable medical devices (IMD) opens opportunities for dynamic real-world monitoring to assess associated risks related to implanted materials. Due to population ageing and expanding demands, total hip, knee, and shoulder arthroplasties are increasing. Automating the collection and analysis of orthopedic device features could benefit physicians and public health policies enabling early issue detection, IMD monitoring and patient safety assessment. A machine learning tool using natural language processing (NLP) was developed for the automated extraction of operation information from medical reports in orthopedics. A corpus of 959 orthopaedic operative reports from 5 centres was manually annotated using the Prodigy software® with a strong inter-annotator agreement of 0.80. Data to extract concerned key clinical and procedure information (n= 9) selected by a multidisciplinary group based on the French health authority checklist. Performances parameters of the NLP model estimated an overall strong precision and recall of respectively 97.0 and 96.0 with a F1-score 96.3. Systematic monitoring of orthopedic devices could be ensured by an automated tool, leveraging clinical data warehouses. Traceability of medical devices with implantation modalities will allow detection of implant factors leading to complications. The evidence from real-world data could provide concrete and dynamic insights to surgeons and infectious disease specialists concerning implant follow-up, guiding therapeutic decision-making, and informing public health policymakers. The tool will be applied on clinical data warehouses to automate information extraction and presentation, providing feedback on mandatory information completion and

[1] Corresponding Author: Ansoborlo Marie; E-mail: m.ansoborlo@chru-tours.fr.

contents of operative reports to support improvements, and thereafter implant research projects.

Keywords. Natural Language processing, Text classification, Electronic Health Records, Operative reports, Orthopaedic material surveillance

1. Introduction

In the field of implantable medical devices (IMD), where the surveillance of surgical site infections is already semi-automated in hip and knee arthroplasty in France (ISO-ORTHO indicator) [1–3], the use of health data opens up possibilities for dynamic real-world monitoring of material, surgical techniques, and patient profiles [4,5]. Total hip, knee and shoulder arthroplasties are increasing due to population ageing and expanding demands [6–8]. Given arthroplasty is common and also a costly procedure, documenting outcomes after joint replacement remains a priority for patient safety and implant performance [9,10]; materio-vigilance is a key target for epidemiologic surveillance [3,11]. Artificial intelligence (AI) tools such as machine learning builds from traditional statistical methods has drawn significant interest in healthcare epidemiology [12–14]. The utilization of big data from electronic health records to automate surveillance of IMDs could be promoted as an innovative approach for routine surveillance. This is particularly relevant given the low frequency of implant re-interventions and the recent updated requirements in material-vigilance [12,13]. Indeed, automating the collection and analysis of implant features and surgical techniques directly from the operative reports could benefit physicians and public health policies, enabling early issue detection and decision support. Researchers have currently access to large electronic healthcare databases due to the implementation of clinical data warehouses (CDWs). In France, the West network of CDWs represents one of the first CDW models (eHOP®), deployed across the 5 university hospitals of the Grand Ouest region of France [15–19], corresponding to the largest European network (three French administrative regions, 12 million individuals). This network is supported by an interregional hub, the Ouest Data Hub - ODH. These CDWs contain clinico-biological, pharmaceutical and claim data. They are structured to be directly usable for research, going beyond the limits of existing registries [15,16,20] by including dynamic real-world data with thinner granulometry.

The HACRO-HUGORTHO project led by the interregional network of orthopaedic surgeons (HUGORTHO®) aimed to propose a robust supervised machine learning tool using natural language processing (NLP) to automate extraction of mandatory information concerning the surgery and patient's condition from operative reports, participating in the construction of the new material-vigilance era. The ultimate goal was to assess the information completion and provide feedback to surgeons to support reporting and quality improvements.

2. Methods

An algorithm based on the claim data source was deployed in the CDWs of all the participating orthopaedic departments of the *Grand Ouest* region to identify the hospital stays involving a hip, knee or shoulder arthroplasty. The claim data source includes medico-administrative data from each hospital stay for which coding is mandatory:

diagnoses based on the International Statistical Classification of Diseases, 10th edition (ICD-10) and medical procedures based on the French Current Procedural Terminology- (CPT; *Classification commune des actes médicaux*). Hence, the identification of the patients of interest relied on procedure codes from the French CPT. The inclusion period encompassed patients undergoing surgery between 2019, January 1st and 2021, December 31st. To constitute a multicentre corpus, 1,000 orthopaedic operative reports were extracted from the five local CDWs, from the electronic health record source to consider potential report discrepancies expected to be around 20% (26–28). Eligibility of the operative reports was assessed based on three criteria: i) an operative report (Loinc: LN:34874-8), ii) first joint replacement on a native joint, iii) without any comorbidities concerning cancer nor infection. Reports were annotated with the Prodigy® tool for nine labels about the patient (diagnosis, joint side) and the surgery (antibiotic prophylaxis, anaesthesia, surgical access incision, complications, cutaneous closure, drainage, joint immobilization). The labels were selected based on the multidisciplinary expertise, both in orthopaedics with HUGORTHO® and in medical informatics with the clinical data centres, in relation with the updated expectations of the European materio-vigilance regulation and expert recommendations and consistently among joint location [12]. Annotations were realized by one orthopaedic surgeon and data centre physicians. Inter-annotator agreement and agreement between annotators and the chosen model were estimated by the Cohen's kappa coefficient κ. Expected patterns were defined precisely to increase the homogeneity intra-label (i.e. few distinct patterns for one label) in one hand, and the heterogeneity inter-labels (i.e. high variability between distinct labels) in the other hand. Eventually, a strong inter-annotator agreement was estimated (κ 0.80) along with a agreement annotators/model about (κ 0.79). A multiclassification model based on neural networks (Spacy®) was trained and tested. The pre-training Spacy® corpus was based on four French text sources concerning political, newspapers, medical and general articles (French Europarl, Newspapers *Est Républicain*, European Medicines and Wikipedia). Our corpus of operative reports was split up in 80% for the train/validation set and 20% for the test. The model performances were estimated with an average F1- score for labels and joints [14,16]. The retained model was the model with the best F1-score.

3. Results

A multicentre corpus of 1,000 orthopaedic reports was available and extracted from the electronic health record database of participating CDWs. After a manual check, a total of 959 documents were included: 317 for the shoulder, 290 for the knee and 352 for the hip. The two main causes of exclusion were a non-primary arthroplasty and a medical history of cancer or infection. The most frequent labels found were the diagnosis (97.9%) and the surgical incision (95.6%). The rarest labels were the complication (21.4%) and the antibiotic prophylaxis (54.5%). The model was trained on a corpus of 767 operative reports and performance parameters were estimated on the test corpus (192 operative reports). F1-score varied between 86.9 and 98.5 depending on the label, Table 1.

Table 1. Performances parameters of the natural language processing classifier by label

Label	Precision	Recall	F1-score
Overall	*95.3*	*97.4*	*96.3*
Anesthesia	97.1	99.5	98.5
Antibiotic prophylaxis	98.2	97.4	97.8

Surgical access incision	99.0	93.0	97.6
Diagnosis	99.0	96.5	97.2
Joint immobilization	96.4	97.0	96.4
Drainage	98.0	93.3	95.3
Surgical access incision	98.3	96.7	94.0
Complications	95.8	92.0	91.4
Joint side	95.3	79.8	86.9

4. Discussion and Conclusions

This study proposed a robust supervised machine learning tool using natural language processing (NLP) to automate classification of operative reports, based on the presence of surgery and patient key information. Indeed, the performance parameters were higher than expected with the pre trained SpaCy® model F1-score at 96.3 overall labels and joints. This project falls within the scope of studies on the use and safety of healthcare products in real-world settings and the development of tools for pharmaco-epidemiology. The tool could now be applied on operative reports to automate reporting concerning the completeness of mandatory information in the operative reports. Feedback to orthopaedists surgeons and infectious disease specialists about implant follow-up, guiding them in therapeutic decision-making, and informing public health policymakers could implement future research based on information extracted from the tool. This pre-intervention part of the HACRO-HUGORTHO study, developing a machine learning tool of automated extraction and analysis of arthroplasty reports, highlighted the completion discrepancies in the operative reports, driving the implementation of a common convergent operative report frame to support quality improvement. Indeed, presented to the surgeons, the overall completeness of the operative reports concerning the nine labels gave a starting point at the quality management program within the HUGORTHO® network, and the West data centre network. The study presented some limits. The external validity of the model evaluated may be affected by the source of the report corpus, selected from university hospitals, where operative reports are already semi-structured with templates presenting key words. Our NLP tool may therefore be more transposable to similar settings. Despite the common NLP constraints (scarcity of specific textual data resources and time-consuming computational process), a unique model was developed over all the three joints and nine labels. Additionally, our tool was not configured to determine the eligibility of an operative report within the electronic health record data source, which could result in predicting labels for documents that are not relevant to the study. Trackability of medical devices with implantation modalities will allow detection of implant factors leading to complications, even rare such as surgical site infections, enhancing future research projects and hypotheses to study.

In conclusion, automated surveillance tools will eventually ensure systematic monitoring of patients and their implanted medical devices, regardless of the healthcare facility where surgery was performed. These tools can provide timely and dynamic responses to orthopaedists, cardiologists, and infectious disease specialists regarding implant follow-up, aiding their therapeutic decisions. They will also guide public health decision-makers in shaping national health policies for medical device surveillance. The role of big data in reforming vigilance in Europe must be considered. Our collaborative team research, especially the feasibility study for an e-cohort for the surveillance of implants in orthopaedics following the implementation of the semi-automated ISO-ORTHO indicator by the national authority of health [2,3,11] demonstrated the interest of artificial intelligence tools in healthcare for efficient and cost-effective automated surveillance, in compliance with recent regulations [12].

References

[1] Grammatico-Guillon L, Miliani K, Banaei-Bouchareb L, Solomiac A, Sambour J, May-Michelangeli L, et al. A computerized indicator for surgical site infection (SSI) assessment after total hip or total knee replacement: The French ISO-ORTHO indicator. Infect Control Hosp Epidemiol. 2022;43:1171-8.

[2] Grammatico-Guillon L, et al. Validation of the first computerized indicator for orthopaedic surgical site infections in France: ISO-ORTHO. Antimicrob Resist Infect Control. 2023;12:44.

[3] Grammatico-Guillon L, Baron S, Gaborit C, Rusch E, Astagneau P. Quality assessment of hospital discharge database for routine surveillance of hip and knee arthroplasty-related infections. Infect Control Hosp Epidemiol Off J Soc Hosp Epidemiol Am. juin 2014;35(6):646-51.

[4] Dhalluin T, Fakhiri S, et al. Role of real-world digital data for orthopedic implant automated surveillance: a systematic review. Expert Rev Med Devices. août 2021;18(8):799-810.

[5] Dhalluin T, et al. Pilot Study of an e-Cohort to Monitor Adverse Event for Patient with Hip Prostheses from Clinical Data Warehouse. Stud Health Technol Inform. 18 nov 2021;287:45-9.

[6] Grammatico-Guillon L, et al. Surgical site infection after primary hip and knee arthroplasty: a cohort study using a hospital database. Infect Control Hosp Epidemiol. oct 2015;36(10):1198-207.

[7] Lemaignen A, Grammatico-Guillon L, et al. Computerized registry as a potential tool for surveillance and management of complex bone and joint infections in France: French registry of complex bone and joint infections. Bone Jt Res. oct 2020;9(10):635-44.

[8] Nemes S, Gordon M, Rogmark C, Rolfson O. Projections of total hip replacement in Sweden from 2013 to 2030. Acta Orthop. juin 2014;85(3):238-43.

[9] Marmor S, Farman T. Causes de procédures médicolégales après prothèse totale de hanche. Rev Chir Orthopédique Traumatol. 1 nov 2011;97(7):752-7.

[10] Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic Burden of Periprosthetic Joint Infection in the United States. J Arthroplasty. sept 2012;27(8, Supplement):61-65.e1.

[11] Grammatico-Guillon L, et al. A computerized indicator for surgical site infection (SSI) assessment after total hip or total knee replacement:The French ISO-ORTHO indicator. Infect Control Hosp Epid. 2021;1-8.

[12] Niemiec E. Will the EU Medical Device Regulation help to improve the safety and performance of medical AI devices? Digit Health. 2022;8:20552076221089079.

[13] van Mourik MSM, et al. PRAISE: providing a roadmap for automated infection surveillance in Europe. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. juill 2021;27 Suppl 1:S3-19.

[14] Hamilton AJ, et al. Machine learning and artificial intelligence: applications in healthcare epidemiology. Antimicrob Steward Healthc Epidemiol. 2021;1(1):e28.

[15] Cuggia M, Bayat S, Garcelon N, Sanders L, Rouget F, Coursin A, et al. A full-text information retrieval system for an epidemiological registry. Stud Health Technol Inform. 2010;160(Pt 1):491-5.

[16] Bories M, Bouzillé G, Cuggia M, Le Corre P. Drug-Drug Interactions with Oral Anticoagulants as Potentially Inappropriate Medications: Prevalence and Outcomes in Elderly Patients in Primary Care and Hospital Settings. Pharmaceutics. 5 juill 2022;14(7):1410.

[17] Bannay A, et al. Leveraging National Claims and Hospital Big Data: Cohort Study on a Statin-Drug Interaction Use Case. JMIR Med Inform. 13 déc 2021;9(12):e29286.

[18] Lemordant P, et al. How to Optimize Connection Between PACS and Clinical Data Warehouse: A Web Service Approach Based on Full Metadata Integration. In: Otero P, Scott P, Martin SZ, Huesing E, éditeurs. Studies in Health Technology and Informatics [Internet]. IOS Press; 2022 [cité 8 août 2023]. Disponible sur: https://ebooks.iospress.nl/doi/10.3233/SHTI220025

[19] Madec J, et al. eHOP Clinical Data Warehouse: From a Prototype to the Creation of an Inter-Regional Clinical Data Centers Network. Stud Health Technol Inform. 21 août 2019;264:1536-7.

[20] Cuggia M, Combes S. The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. Yearb Med Inform. août 2019;28(1):195-202.