**FACULTÉ DE MÉDECINE ET PHARMACIE DE ROUEN**

**ANNÉE 2023**

**MEMOIRE EN MÉDECINE**

**PAR**

**Francesco MONTI**

**NÉ LE 27 JUIN 1990 À BAGNO A RIPOLI (FI, ITALIE)**

**PRÉSENTÉE ET SOUTENUE PUBLIQUEMENT LE 26 SEPTEMBRE 2023**

**EFFORTLESS IDENTIFICATION OF SURGICAL SITE INFECTIONS: EMPOWERING CLINICAL DATA WAREHOUSES**

**Président du jury:** M. le Professeur Jacques BENICHOU

**Directeur de mémoire:** M. Julien GROSJEAN, PhD

**Membre du jury:** M. le Professeur Stefan DARMONI

**Membre du jury:** Mme le Professeur Marie-Pierre TAVOLACCI

Summary

[Abstract 3](#_Toc145332234)

[1. Introduction 4](#_Toc145332235)

[2. Methods 5](#_Toc145332236)

[2.1. Population 5](#_Toc145332237)

[Inclusion criteria 5](#_Toc145332238)

[Exclusion criteria 6](#_Toc145332239)

[2.2. EDSaN – RUH’s Clinical data warehouse 6](#_Toc145332240)

[2.3. Algorithm evaluation 8](#_Toc145332241)

[3. Results 9](#_Toc145332242)

[Time spent on the reviewing process 11](#_Toc145332243)

[4. Discussion 11](#_Toc145332244)

[Strengths 11](#_Toc145332245)

[Limitations 13](#_Toc145332246)

[5. Conclusions 14](#_Toc145332247)

[References 16](#_Toc145332248)

[Abbreviations 18](#_Toc145332249)

[Appendix 19](#_Toc145332250)

[Workflow 19](#_Toc145332251)

[Identification of surgery events 21](#_Toc145332252)

[Identification of infective events 26](#_Toc145332253)

[Characteristics of the 300 random selected cases 29](#_Toc145332254)

[Metrics Legend 29](#_Toc145332255)

[Prevalence Metrics 29](#_Toc145332256)

[Basic Performance Metrics 30](#_Toc145332257)

[Predictive Values 30](#_Toc145332258)

[Likelihood Ratios 30](#_Toc145332259)

[Additional Metrics 30](#_Toc145332260)

[Overall Performance 30](#_Toc145332261)

# Abstract

**Background**: Surgical Site Infections (SSIs) are a prevalent and costly issue in healthcare, often leading to extended hospital stays and increased morbidity. Traditional methods for detecting SSIs are manual, cumbersome, and prone to human error. This study aims to assess the feasibility and effectiveness of an automated surveillance system using Electronic Health Record (EHR) data for the targeted detection of SSIs following spinal surgeries.

**Methods**: Leveraging the capabilities of Entrepôt des Données de Santé Normand (EDSaN), a specialized Clinical Data Warehouse at Rouen’s University Hospital, we identified a cohort of patients who underwent spinal surgeries between January 1, 2020, and December 31, 2020. These patients were subsequently monitored for the development of SSIs. The algorithm employed multiple methodologies, including Natural Language Processing queries and standardized ICD-10 codes, for comprehensive SSI identification. A manual review of a random sample of 300 cases was conducted to validate the algorithm's performance, employing metrics such as sensitivity, specificity, positive predictive value, and negative predictive value.

**Results**: Out of 652 patients who underwent spinal surgery, the algorithm flagged 79 for readmission due to postoperative SSIs, equating to a prevalence rate of 12.11%. In terms of algorithmic effectiveness, sensitivity was found to be 0.82, and specificity was 0.98. Importantly, the enhanced workflow reduced the average time required for manual case review to 5.75 minutes, signifying notable gains in operational efficiency.

**Conclusions**: Our findings indicate that an automated surveillance system using EHR data can offer a highly effective and efficient method for SSI detection. This system not only demonstrates robust diagnostic accuracy but also shows promise in substantially reducing the manual labor involved in surveillance activities. Furthermore, the adaptability of this approach suggests its potential applicability in the monitoring of other types of healthcare-associated infections.

**Keywords**: Surgical Site Infections, Electronic Health Records, Automated Surveillance, Clinical Data Warehouse, Healthcare-Associated Infections, Operational Efficiency, Natural Language Processing

# Introduction

As Electronic Health Record (EHR) systems gain ubiquity, the volume of electronic clinical data continues to burgeon. This proliferation has captivated researchers, healthcare administrators, and clinicians alike, fueling interest in the secondary utilization of EHR data to augment clinical acumen and optimize patient care. Among the myriad applications of EHR data, the targeted detection of specific outcomes and adverse conditions—such as Surgical Site Infections (SSIs)—emerges as a compelling avenue for exploration.(1–4)

Surgical site infections (SSIs) are among the most frequently cited hospital-acquired infection and lead to significant morbidity, prolonged hospitalization, increased medical costs, and overall compromised patient outcomes (5). In particular, SSIs following spine surgery can be devastating, requiring surgical debridement(s) and prolonged intravenous antibiotics, and at times, leading to significant long-term disability. Multiple studies have outlined general risk factors for SSIs, and both the medical condition of the patient and the complexity of the surgical procedure are contributing factors. (6,7)

SSIs are defined as infections that occur at or near the surgical incision within 30 days of the operation, or within one year if an implant is in place (8). They are classified based on the depth and severity into:

* Superficial Incisional: Infection involving only the skin and subcutaneous tissue of the incision.
* Deep Incisional: Infection involving deep tissues, such as fascia and muscle layers.
* Organ/Space: Infection involving any part of the anatomy other than the incision, which was opened or manipulated during the operation.

The French Health Authority ardently advocates for the routine surveillance of SSIs, situating it within the broader framework of risk management for Healthcare-Associated Infections (HAIs) (9). Current SSI detection practices, largely reliant on manual surveillance by local hospital hygienists or nurses, are fraught with inconsistencies and limitations—ranging from human error to staff turnover and training gaps. This manual approach is not only resource-intensive than it could be but also diverts critical resources away from the conceptualization and monitoring of preventive strategies. The (semi)automation of SSI detection via EHRs could offers a more standardized, efficient, and comprehensive modus operandi or at the very least, enhance current workflow.

In recent years, Clinical Data Warehouses (CDWs) have emerged as indispensable assets within hospital settings, facilitating the extraction of actionable insights from both structured and unstructured data. These tools predominantly employ Natural Language Processing (NLP) algorithms to sift through clinical narratives for a variety of applications, such as identifying eligible patients for clinical trials and targeted research endeavours.

This study aims to evaluate the effectiveness of leveraging Rouen’s University Hospital's Clinical Data Warehouse, known as Entrepôt des Données de Santé Normand (EDSaN), in identifying SSIs in spinal surgeries. The ultimate goal is to enhance the current surveillance activities and to reduce the operational challenges tied to current manual methods. It is important to note that this system is designed to complement human labor, aiming to provide healthcare professionals with an interface for rapid access to relevant patient information and flagging critical elements that warrant review.

# Methods

## Population

### Inclusion criteria

The study focuses on patients who underwent any form of spinal surgery at RUH between January 1, 2020, and December 31, 2020, and experienced SSIs postoperatively. Surgical procedures are identified using the Classification Commune des Actes Médicaux (CCAM) list, developed in collaboration with the Medical Information Department (MID). This methodology ensures that only patients operated on at RUH are included, thereby providing a robust and consistent approach for targeting the relevant population.

For the identified patient cohort, potential SSIs were ascertained by examining all subsequent hospitalizations post-surgery. This also includes instances where the SSI occurred during the same hospital stay as the surgery. The identification of infections was achieved through a combination of Natural Language Processing (NLP) queries, ICD-10 codes, and/or CCAM acts. These multiple methodologies serve to compensate for the intrinsic limitations of each approach. For example, while NLP queries can capture specific clinical details in medical reports, ICD-10 and CCAM codes offer standardization that facilitates large-scale analysis.

It is important to note that while the date of the surgical intervention is explicitly recorded in the database, the exact date of infection onset is not consistently available and should generally slightly precede a readmission. In the absence of reliable data from clinical reports, the date of hospital readmission will serve as a proxy for the infection event date. However, this methodological choice may introduce variability in adhering to the temporal criteria established for the definition of SSIs. To mitigate this limitation, the time filter between the infection date and the surgical intervention date may include a certain tolerance.

### Exclusion criteria

Patients who do not meet the definition of an SSI as described are excluded from the cohort (8).

## EDSaN – RUH’s Clinical data warehouse

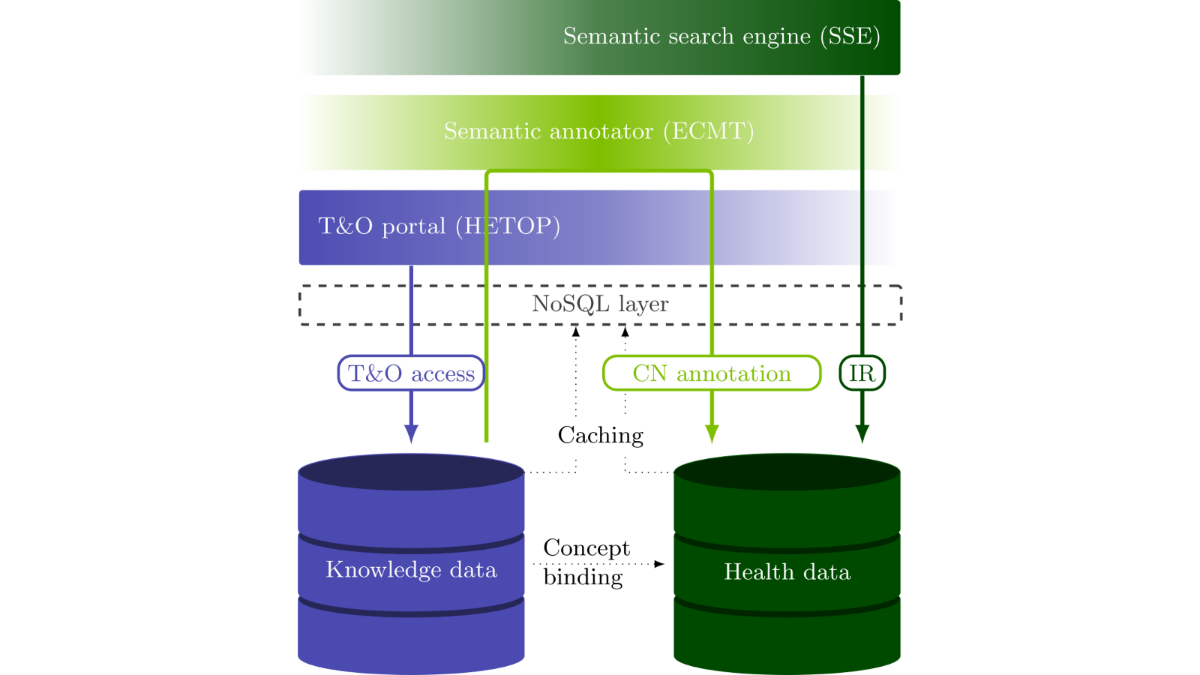
All the Rouen’s University Hospital (RUH’s) digitalized documents are available through EDSaN (10), the local Clinical Data Warehouse (CDW), gathering together the health information of the about 2 million patients who visited the hospital since 1998 (11). The SHDW currently focuses on clinical data and, more broadly, on health data according to a patient-centered strategy. In addition to the structured patient data, the different data pertaining to multiple admissions and events at RUH are collected (eg, diagnoses, biology, procedures, and movements). Unstructured data embedded in electronic health records (EHR) (mostly narrative reports) are necessary to solve trial eligibility criteria in 59% to 95% of clinical studies (12). Indeed, a wide range of crucial healthcare data is commonly found within unstructured clinical narratives as they allow flexibility of expression such as doubts, negations, or diagnostic hypotheses and complex representation of diseases, clinical examination or patient history. The 23 million clinical documents in French of RUH consequently play a strategic role in the context of the CDW.

The SHDW enables the semantic retrieval of health data in French based on several terminologies and ontologies (T&Os) and consequently relies on two datasets: a domain knowledge database and a health database maintaining clinical and patient data.

The reference-controlled vocabularies (ie, reference management domain) providing the knowledge database are notably widely collected and maintained through the cross-terminological portal HeTOP. All this data integrates into a modular architecture that can be queried via a graphical user interface (GUI). The functionalities of the semantic health data warehouse (SHDW) are ensured by the collaboration of three distinct layers, where each layer consumes data from the above layers (see Figure 1):

1. the cross-terminological portal HeTOP (13)
2. semantic annotator ECMT (14,15)
3. semantic search engine (SSE) (16–18).

Figure 1: Functional architecture of the semantic health data warehouse



Once a population has been selected it can be exposed via the EDSaN Consult, an intuitive interface designed to provide healthcare professionals with rapid access to patient information. This interface allows users to efficiently review all digitized information related to identified events and categorize patients into various groups such as true positives, false positives, and more, or create custom categories. Notably, elements that triggered matches in the queries are flagged for a quick and easy identification.

Furthermore, EDSaN consult offers flexibility in the display of information. Users can choose to view the entire patient history, only elements that matched the queries or all elements related to the events found by the queries. Additionally, dataset can be exported for further analysis.

The system is designed to enhance the existing workflow and complement human oversight. By providing rapid access to pertinent patient data and highlighting critical elements retrieved by the queries, EDSaN consult helps to streamline research or surveillance processes and reduce the operational burden. This is something that is already and being done routinely for research purposes and, for example, to identify patients at risk of osteoporotic fracture aiming to integrate them into a dedicated care program. As of September 2023, EDSaN has been exploited in over 400 use cases/research projects and numerous consequent publications (19–21).

## Algorithm evaluation

In the context of information retrieval, two key metrics are commonly employed to evaluate the performance of search algorithms: *precision* and *recall*.

*Precision* is defined as the proportion of accurately identified instances relative to the total instances returned by the algorithm. Conversely, *recall* quantifies the proportion of accurately identified instances relative to the total instances that should have been identified. Employed individually, these metrics offer limited utility. For example, achieving perfect recall is feasible by indiscriminately retrieving all items, both relevant and irrelevant. Similarly, near-perfect precision can be attained by selectively retrieving a minuscule subset of highly probable items. Consequently, these metrics are often either compared at a fixed level (e.g., precision at a recall level of 0.75) or amalgamated into a composite measure, such as the F-measure or the Matthews correlation coefficient.

A significant limitation of this study is the absence of a manually validated, comprehensive list of patients with SSIs for each year, which would serve as a gold standard for performance evaluation. This absence precludes a definitive benchmark for our search algorithm. Nevertheless, the exhaustive nature of the surgical procedures list used for population selection suggests that our results should closely approximate actual conditions. Readers should be cognizant of this potential bias.

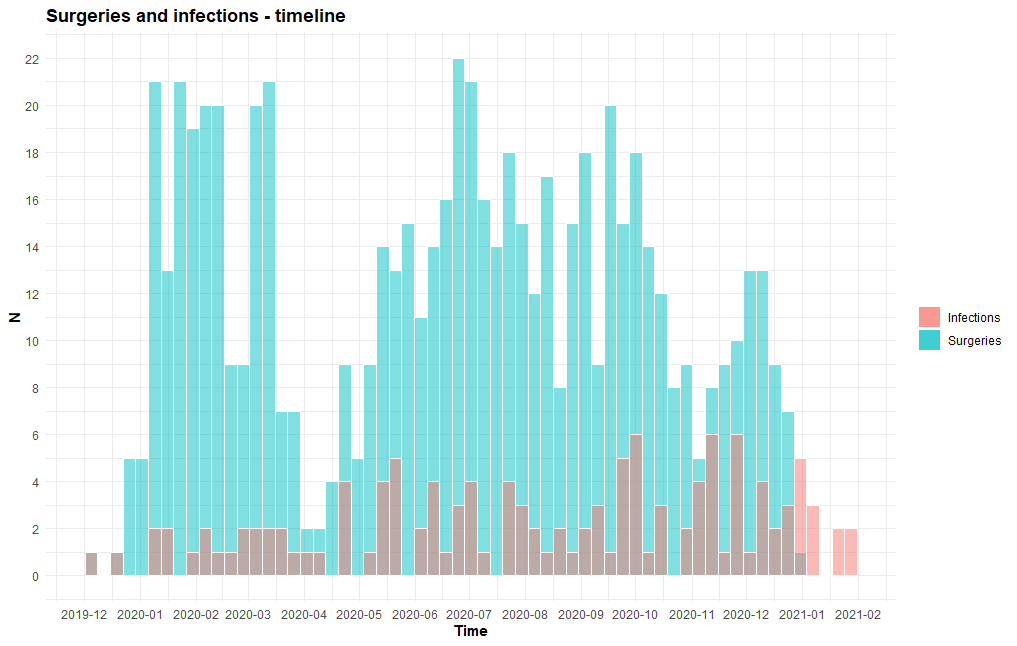
To mitigate this limitation, 300 patients were randomly selected from the 652 identified and subjected to manual review, with several standard epidemiological metrics calculated. This sample size should be sufficiently representative to permit result extrapolation. Additionally, the time expended on this manual review will be documented to gauge potential efficiency gains and workload reductions for the Hygiene Department. It is worth noting that this will be an approximate estimate, as the Hygiene Department has not previously assessed the average time dedicated to individual SSI cases.

Despite these limitations, this study aims to offer valuable insights into the capabilities of the local Health Data Warehouse (HDW) and to advocate for the adoption of this solution for similar use-cases. For additional details on the inclusion process and workflow, please consult the Appendix.

# Results

In the year 2020, the search algorithm identified 652 patients who had undergone spinal surgery. Of these, 79 were readmitted to the hospital due to postoperative infections, resulting in a Surgical Site Infections (SSIs) prevalence rate of 12.11%. Below, in Figure 2, you can see the timeline of interventions and infections events.

Figure 2: timeline of surgical and infective events



Among the 79 infections identified, accounting for 96 readmissions, 77 met both the timing and material implant criteria as stipulated in the SSI definition. On average, infections manifested after a delay of 19.1 days, with a range spanning from a few days for infections occurring during the same hospital stay, to as long as 111 days. The primary measures of dispersion are summarized in the Table 1 below:

Table 1: Dispersion measures - infection onset delay.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | Mean | SD | Min | Max | Range |
| Hospital stays infection + | 96 | 19.21 | 26.48 | 0 | 111 | 111 |

In line with the algorithm's findings, 35 of the 77 infections, constituting 45.45%, manifested immediately following the surgical intervention, occurring during the same hospital stay.

A representative sample of 300 cases was randomly selected for manual review. For additional insights into the temporal distribution and dispersion metrics of these randomly selected infections (which looks statistically consistent with the overall cohort), please consult the Appendix.

The results of the classification task are summarized in Table 2 and Table 3.

Table 2 Contingency table summarizing reviewing results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Review + | Review - | Total |
| Algo + | 37 | 6 | 43 |
| Algo - | 8 | 249 | 257 |
| Total | 45 | 255 | 300 |

Table 3 Evaluation metrics.

|  |  |  |
| --- | --- | --- |
| Metric | Estimate | 95% CI |
| Apparent prevalence | 0.14 | 0.11, 0.19 |
| True prevalence | 0.15 | 0.11, 0.20 |
| Sensitivity (recall) | 0.82 | 0.68, 0.92 |
| Specificity | 0.98 | 0.95, 0.99 |
| Positive predictive value (precision) | 0.86 | 0.72, 0.95 |
| Negative predictive value | 0.97 | 0.94, 0.99 |
| F1-score | 0.841 |  |
| Matthews coefficient correlation | 0.814 |  |
| Positive likelihood ratio | 34.94 | 15.67, 77.95 |
| Negative likelihood ratio | 0.18 | 0.10, 0.34 |
| False T+ for True D- | 0.02 | 0.01, 0.05 |
| False T- for True D+ | 0.18 | 0.08, 0.32 |
| False T+ proportion for T+ | 0.14 | 0.05, 0.28 |
| False T- proportion for T- | 0.03 | 0.01, 0.06 |
| Correctly classified proportion | 0.95 | 0.92, 0.97 |

## Time spent on the reviewing process

The time spent on the reviewing process through EDSaN Consult has been assessed and the average time per case was found to be 5.75 minutes, spanning from a minimum of 1 minute to a maximum of 17.6 minutes. Primary dispersion measures are resumed in the Table 4 below.

Table 4: reviewing time primary dispersion measures.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | Mean | Sd | Min | Max | Range |
| Reviewing time | 300 | 5.75 | 5.57 | 1 | 17.65 | 16.65 |

# Discussion

The possibility of implementing automated detection of Surgical Site Infections (SSIs) represents a thrilling option in healthcare surveillance. Its efficiency, accuracy, and flexibility have the potential to revolutionize the way we approach patient outcomes and healthcare practices.

## Strengths

While the algorithm's specificity of 98% is highly commendable, its sensitivity of 82% may initially appear less impressive. However, performance metrics such as the F1-Score and the Matthews Correlation Coefficient (MCC) provide a more nuanced evaluation. The F1 Score of 0.841 and MCC of 0.814 suggest that the algorithm strikes a good balance between precision and recall and shows a strong correlation between observed and predicted classifications. This speaks to the algorithm's robustness and reliability.

The algorithm is not intended to replace manual validation but to flag potential SSIs for further review. In this light, an 82% sensitivity rate is significant. It means that the algorithm successfully identifies a large majority of actual SSIs, which can then be manually confirmed in a fraction of the time compared to traditional methods. The high specificity further enhances the algorithm's utility by minimizing the number of false positives, thereby making the manual review process more efficient. In a setting where a nurse from the Hygiene Department currently dedicates only four hours a week to SSI surveillance, this efficiency is invaluable. Although we do not have exact data on the time currently spent per case using traditional surveillance methods, the algorithm flagged 652 cases in 2020, and a manual review of 300 of these cases took an average of 6 minutes each. Given this, it is reasonable to hypothesize that the algorithm could represent a very substantial timesaving, as it offers several efficiency gains over traditional methods. It eliminates the need for physical visits to surgery units, removes the dependency on others healthcare professionals availability for dialogue, and allows for more flexible time allocation for SSI surveillance. These factors not only streamline the surveillance process but also significantly enhance the capabilities of the surveillance staff, allowing them to focus on more complex tasks requiring human expertise.

We can go a bit further and calculate rough estimates of the possible efficiency gains. Currently, a nurse in the Hygiene Department dedicates 4 hours per week to SSI surveillance, totaling approximately 10,560 minutes per year, considering 2 months of vacation time. In contrast, the algorithm identified 652 potential SSI cases in 2020, requiring an estimated 3,912 minutes for manual review (based on the 6 minutes on average it took to review the 300 randomly selected cases). This represents a huge *potential* timesaving of around 6,648 minutes per year, or approximately 63% of the time currently allocated for this task. These figures suggest that the algorithm could substantially enhance the efficiency of SSI surveillance, allowing healthcare providers to allocate their time more effectively.

In France, the cost of employees is among the highest in Europe, often adding an additional 35% due to social security contributions, taxes, and other benefits (22). If we suppose a base salary of €2,000 per month, the total annual cost to the hospital would be approximately €32,400. Given that the current manual method of SSI surveillance occupies about 10% of a nurse's annual working time, we can estimate the potential financial savings.

By implementing the algorithm, we estimate a potential timesaving of 63%, which translates to around 111 hours saved annually per nurse. In monetary terms, this could result in an annual saving of approximately €2,041.2 per nurse dedicated to SSI surveillance.

This does not even account for the qualitative benefits like increased accuracy and the ability to reallocate nursing time to other critical tasks. It is worth noting that these are conservative estimates and the actual savings it's likely to be higher. Nevertheless, the exact timesaving benefits remain to be quantified.

Beyond spinal surgeries, the algorithm has the potential to be adapted for other surgical types and complications, making it a versatile and scalable solution for comprehensive SSI surveillance providing surgeons with a much-needed closer feedback loop.

## Limitations

An area that requires attention and improvement is the sensitivity rate. As is often the case, information retrieval in the hospital setting is hampered by vague language in clinical notes, which poses a significant challenge in achieving higher sensitivity. This issue is twofold:

* Vague terminology: Often, clinicians inadvertently use vague or non-standard terminology without considering its impact on automated interpretation. This can leads to false negatives, where the algorithm fails to identify a true case of SSI.
* Reluctance to document complications: based on the at times extremely vague language employed there seems to be a reluctance among surgeons to explicitly document complications, possibly due to concerns about professional reputation or medico-legal implications. This lack of clear documentation further contributes to the algorithm's inability to identify true positive cases.

This experience underscores the imperative for direct and open communication between the Medical Informatics Department and clinicians. Such collaboration is not only vital for understanding the nuances of clinical language but also for fostering a culture of transparency and accountability. By working closely with clinicians to refine the algorithm, we can aim to improve its sensitivity, thereby enhancing the effectiveness and accuracy of automated SSI surveillance systems. In exchange, the algorithm would provide an easy avenue for clinicians to receive feedback on their practices. By automating the detection and reporting of SSIs, it creates a more transparent and immediate feedback loop, which could be instrumental in improving surgical procedures and post-operative care.

The absence of a comprehensive gold standard for performance evaluation remains a significant limitation. However, the high specificity and sensitivity indicated by the metrics suggest that the algorithm is surely a step in the right direction.

While this is not necessarily an algorithm’s specific limitation, it does not address the issue of the "Unknown Unknowns": If patients do not return to the hospital for their infections, these cases will never be captured. This limitation raises questions about the true prevalence of SSIs and highlight once again how instrumental it would be to have a bridge between hospital healthcare and general practitioners (23).

The algorithm makes certain simplifications, such as equating the date of surgery with the first hospitalization and the date of infection with readmission. While practical, these assumptions could affect the algorithm's ability to accurately capture the timing of infections.

The study does not address the potential conflict of interest that arises from the manual review of retrieved instances being conducted by the same person that developed the algorithm.

Lastly, the performance of the algorithm may have been impacted by incomplete digitalization of hospital information, as not all relevant data could be readily accessible in electronic format.

# Conclusions

The development of this algorithm has been an enlightening journey, highlighting the critical role of interdisciplinary collaboration between data scientists, clinicians, and public health experts. The efficiency gains are not merely numerical but translate into tangible benefits for healthcare providers. By automating a significant portion of the SSI surveillance process, the algorithm frees up valuable time, allowing healthcare providers to focus on other critical aspects of patient care that cannot be automated, such as nuanced clinical judgment and patient interaction.

The algorithm's high specificity and sensitivity are promising, but there is room for improvement, particularly in increasing the sensitivity rate. Future iterations could explore the integration of more complex machine learning models or the inclusion of additional variables like patient comorbidities or surgical techniques, which could be indicative of SSIs.

Beyond spinal surgeries, the algorithm could be adapted to other surgical types and even other post-operative complications.

For surgeons, the system serves as a feedback tool for continuous quality improvement. While surgeons are keenly aware of individual SSIs, the system would help watchers broaden the specter of the surveillance. A broader epidemiological perspective is invaluable for understanding the prevalence and patterns of SSIs, thereby enabling targeted interventions and quality improvement measures.

The convenience of digital access to EHRs is another significant advantage, eliminating the need for time-consuming physical visits to different units. While we cannot precisely quantify the time saved, the efficiency gains are palpable and align well with the broader goals of healthcare optimization.

Our findings reiterate the importance of open communication channels between the Medical Informatics Department and clinicians. This collaboration is not a one-off requirement but an ongoing process, essential for the iterative refinement of the algorithm and for ensuring its clinical relevance and accuracy

### Future Directions

* Interdisciplinary collaboration: formalizing collaborative frameworks could refine both the algorithm and its clinical applications.
* User experience: usability studies could assess how healthcare providers interact with the system, aiming to make it more user-friendly.
* More advanced machine learning techniques: the current work could serve as a springboard for employing more advanced machine learning models.
* Cost-benefit analysis: comprehensive studies could provide a holistic perspective on both the qualitative and quantitative impacts of the algorithm.
* Clinical adaptability: separate studies could explore the algorithm's applicability in different surgical settings and for other healthcare-associated infections.
* Patient-centered outcomes: research could focus on how automated SSI surveillance impacts patient outcomes directly.
* Data quality: further studies should aim to standardize what constitutes 'good quality' data for automated systems.
* Interregional collaboration: regional partnerships with the other University Hospitals of Normandy (Lille, Caen and Amiens, part of the G4 network together with RUH) could enrich the dataset and offer the possibility to broaden his application.

By exploring these avenues, we not only aim to refine the current algorithm but also to contribute to broader healthcare optimization efforts, potentially revolutionizing how we approach patient care and clinical surveillance.

# References

1. Colborn KL, Bronsert M, Amioka E, Hammermeister K, Henderson WG, Meguid R. Identification of surgical site infections using electronic health record data. Am J Infect Control. 2018 Nov;46(11):1230–5.

2. Campillo-Gimenez B, Garcelon N, Jarno P, Chapplain JM, Cuggia M. Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France. Stud Health Technol Inform. 2013;192:572–5.

3. Grundmeier RW, Xiao R, Ross RK, Ramos MJ, Karavite DJ, Michel JJ, et al. Identifying surgical site infections in electronic health data using predictive models. J Am Med Inform Assoc JAMIA. 2018 Sep 1;25(9):1160–6.

4. Karhade AV, Bongers MER, Groot OQ, Cha TD, Doorly TP, Fogel HA, et al. Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy? Spine J. 2020 Oct 1;20(10):1602–9.

5. Seidelman J, Anderson DJ. Surgical Site Infections. Infect Dis Clin North Am. 2021 Dec;35(4):901–29.

6. Risk Factors for Surgical Site Infections Following Spinal Fusion Procedures: A Case-Control Study | Clinical Infectious Diseases | Oxford Academic [Internet]. [cited 2023 Sep 8]. Available from: https://academic.oup.com/cid/article/53/7/686/421868

7. Nota SPFT, Braun Y, Ring D, Schwab JH. Incidence of surgical site infection after spine surgery: what is the impact of the definition of infection? Clin Orthop. 2015 May;473(5):1612–9.

8. Abdul-Jabbar A, Takemoto S, Weber MH, Hu SS, Mummaneni PV, Deviren V, et al. Surgical site infection in spinal surgery: description of surgical and patient-based risk factors for postoperative infection using administrative claims data. Spine. 2012 Jul 1;37(15):1340–5.

9. Française (SPILF) la S de PI de L, Tropicales (CMIT) C des U de MI et, Pédiatrique (GPIP) G de PI, Réanimation (SFAR) SF d’Anesthésie et de, Traumatologique (SOFCOT) SF de CO et, Hospitalière (SFHH) SF d’Hygiène, et al. Recommandations de pratique clinique. Infections ostéo-articulaires sur matériel (prothèse, implant, ostéo-synthèse). Médecine Mal Infect. 2009 Nov;39(11):815–63.

10. Pressat-Laffouilhère T, Balayé P, Dahamna B, Lelong R, Billey K, Darmoni SJ, et al. Evaluation of Doc’EDS: a French semantic search tool to query health documents from a clinical data warehouse. BMC Med Inform Decis Mak. 2022 Feb 8;22(1):34.

11. Lelong R, Soualmia LF, Grosjean J, Taalba M, Darmoni SJ. Building a Semantic Health Data Warehouse in the Context of Clinical Trials: Development and Usability Study. JMIR Med Inform. 2019 Dec 20;7(4):e13917.

12. Raghavan P, Chen JL, Fosler-Lussier E, Lai AM. How essential are unstructured clinical narratives and information fusion to clinical trial recruitment? AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci. 2014;2014:218–23.

13. Grosjean J, Merabti T, Dahamna B, Kergourlay I, Thirion B, Soualmia LF, et al. Health multi-terminology portal: a semantic added-value for patient safety. Stud Health Technol Inform. 2011;166:129–38.

14. Cabot C, Soualmia LF, Dahamna B, Darmoni SJ. SIBM at CLEF eHealth evaluation lab 2016: Extracting concepts in french medical texts with ECMT and CIMIND. In: Conference and labs of the evaluation forum [Internet]. 2016. Available from: https://api.semanticscholar.org/CorpusID:17094077

15. Soualmia LF, Cabot C, Dahamna B, Darmoni SJ. SIBM at CLEF e-Health evaluation lab 2015. In: Conference and labs of the evaluation forum [Internet]. 2015. Available from: https://api.semanticscholar.org/CorpusID:11176313

16. Lelong R, Cabot C, Soualmia LF. Semantic Search Engine to Query into Electronic Health Records with a Multiple-Layer Query Language.

17. Lelong R, Soualmia L, Dahamna B, Griffon N, Darmoni SJ. Querying EHRs with a Semantic and Entity-Oriented Query Language. Stud Health Technol Inform. 2017;235:121–5.

18. Lelong R, Soualmia LF, Sakji S, Dahamna B, Darmoni S. NoSQL technology in order to support Semantic Health Search Engine. In 2018 [cited 2023 Sep 9]. Available from: https://hal.science/hal-02103574

19. Litovsky J, Hacard F, Tetart F, Boccon-Gibod I, Soria A, Staumont-Salle D, et al. Omalizumab drug survival in chronic urticaria: a retrospective multicentric French study. J Allergy Clin Immunol Pract. 2023 Aug 29;S2213-2198(23)00956-X.

20. Dezoteux F, El Mesbahi S, Tedbirt B, Grosjean J, Gautier S, Lannoy D, et al. Immunomodulatory and/or immunosuppressive drugs should not be stopped prior to skin tests for the assessment of drug allergy. Br J Dermatol. 2022 Apr;186(4):742–4.

21. Léguillon R, Varin R, Pressat-Laffouilhère T, Chenailler C, Chassagne P, Roca F. Clinical Pharmacist Intervention Reduces Potentially Inappropriate Prescriptions in a Geriatric Perioperative Care Unit Dedicated to Hip Fracture. Gerontology. 2023;69(4):386–95.

22. Statistics | Eurostat [Internet]. [cited 2023 Sep 11]. Available from: https://ec.europa.eu/eurostat/databrowser/view/lc\_lci\_lev/default/table?lang=en

23. https://www.p4dp.fr/ [Internet]. [cited 2023 Sep 11]. P4DP. Available from: https://www.p4dp.fr/

# Abbreviations

**CCAM** – Classification Commune des Actes Médicaux

**CDW** – clinical data warehouse

**CI** – Confidence Interval

**DRG** – Diagnosis-Related Group

**ECMT** - Extracteur de Concepts Multi-Terminologique

**EDSaN** – Entrepôt des données de santé de Normandie

**EHR** – electronic health record

**GUI** - graphical user interface

**HDW** – health data warehouse

**HeTOP** - Health Terminology/Ontology Portal

**ICD-10** – International Classification of Diseases, 10th Revision

**IR** - information retrieval

**MCC** - The Matthews Correlation Coefficient

**MDI** - Medical Information Department

**NLP** – Natural Language Processing

**RUH** – Rouen's University Hospital

**SHDW** - semantic health data warehouse

**SSE**- semantic search engine

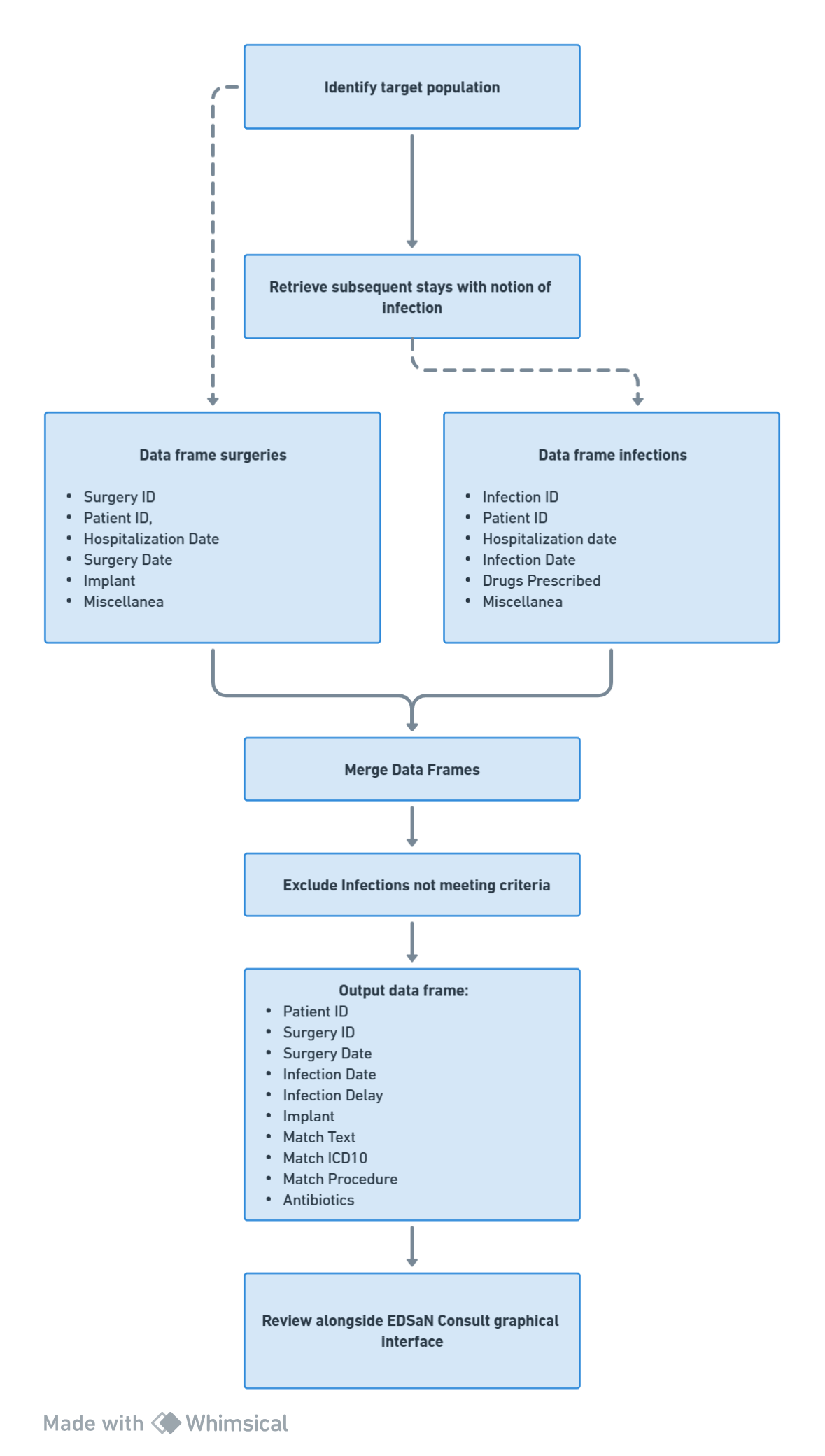
**SSI** – Surgical site infection

**T&Os** - terminologies and ontologies

# Appendix

## Workflow

Figure 3: workflow flowchart

****

Workflow description:

1. identify subjects operated via CCAM codes and extract a data frame with at least the following variables :
   1. Surgery ID
   2. Patient ID
   3. Hospitalization date
   4. Surgery date
   5. Implant (0/1)
   6. Extra variables (demographic data, etc)
2. Retrieve subsequent stays with notion of infection and export a data frame with at least the following variables :
   1. Infection ID
   2. Patient ID
   3. Infection date
   4. Drugs prescribed
3. Merge the two data frames
4. Exclude infections which:
   1. precede the surgery
   2. are more than 30 or 365 days after the date of the intervention based on the presence or absence of implants
   3. are not related to spinal surgeries.
5. Output a data frame with the following variables:
   1. Patient ID
   2. Surgery ID
   3. Surgery date
   4. Infection date
   5. Infection delay (how many days the infection occurred after surgery)
   6. Implant (0/1)
   7. Match text (the phrase containing words that matched the infection query, if any)
   8. Match ICD10 (the ICD-10 codes that matched the infection query, if any)
   9. Match procedure (the procedure.s matched by the infection query, if any)
   10. Antibiotics (has any antibiotic been prescripted)
6. Use the data frame alongside the EDSaN Consult graphical interface to manually review the population

### Identification of surgery events

|  |  |
| --- | --- |
| **Code** | **Label** |
| **LHMH001** | Transcutaneous spondyloplasty of 3 vertebrae with CT guidance |
| **LHMH002** | Single transcutaneous spondyloplasty with X-ray guidance |
| **LHMH003** | Transcutaneous spondyloplasty of 2 vertebrae with CT guidance |
| **LHMH004** | Transcutaneous spondyloplasty of 3 vertebrae with X-ray guidance |
| **LHMH005** | Single transcutaneous spondyloplasty with CT guidance |
| **LHMH006** | Transcutaneous spondyloplasty of 2 vertebrae with X-ray guidance |
| **LDCA002** | Posterior osteosynthesis and/or arthrodesis of the occipitocervical junction without exploration of the canal contents, using a posterior approach |
| **LDCA003** | Posterior osteosynthesis of the occipitocervical junction with exploration of the canal contents, using a posterior approach |
| **LDCA005** | Interlaminar osteosynthesis of the atlas and axis, using a posterior approach |
| **LDCA006** | Transarticular and/or transpedicular osteosynthesis of the atlas and axis, using a posterior approach |
| **LDCA009** | Transpedicular osteosynthesis of the axis, using a posterior approach |
| **LDCA010** | Interlaminar osteosynthesis between two vertebrae of the cervical spine from C2 to C7, using a posterior approach |
| **LFCA001** | Osteosynthesis of the lumbosacral junction with exploration of the canal contents, using a posterior approach |
| **LFCA002** | Osteosynthesis of the lumbosacral junction without exploration of the canal contents, using a posterior approach |
| **LFDA004** | Intercorporeal arthrodesis of the lumbar or lumbosacral spine with posterolateral arthrodesis, using a posterior approach |
| **LFDA012** | Intercorporeal arthrodesis or epiphysiodesis of the lumbar or lumbosacral spine, using a posterior approach |
| **LGCA001** | Osteosynthesis of a fracture of the sacrum, using a posterior approach |
| **LHCA002** | Posterior osteosynthesis of the spine without exploration of the canal contents, using a posterior approach |
| **LHCA010** | Posterior osteosynthesis of the spine with exploration of the canal contents, using a posterior approach |
| **LHCA011** | Osteosynthesis of the spine using an external fixator |
| **LHCA016** | Posterior osteosynthesis of the spine without exploration of the canal contents with arthrodesis, using a posterior approach |
| **LHDA001** | Posterior arthrodesis or epiphysiodesis of the spine without exploration of the canal contents, using a posterior approach |
| **LHDA002** | Intercorporeal arthrodesis of the spine with posterior arthrodesis, using a posterolateral approach |
| **LDCA001** | Bilateral transarticular osteosynthesis of the atlas and axis, by anterior cervicotomy or bilateral anterolateral cervicotomy |
| **LDCA004** | Osteosynthesis of the axis tooth [odontoid process of C2], by anterior or anterolateral cervicotomy |
| **LDCA007** | Osteosynthesis of the cervical vertebral column, by anterior approach with mandibulotomy |
| **LDCA008** | Osteosynthesis of the axis tooth [odontoid process of C2], by intraoral approach |
| **LDCA011** | Osteosynthesis and/or anterior arthrodesis of the vertebral column without exploration of the canal contents, by anterior or anterolateral cervicotomy |
| **LDCA013** | Osteosynthesis of the vertebral column with exploration of the canal contents, by anterior cervicotomy or anterolateral cervicotomy |
| **LDDA001** | Anterior arthrodesis of the occipitocervical junction, using an intraoral approach or anterolateral cervicotomy |
| **LECA001** | Osteosynthesis and/or anterior arthrodesis or epiphysiodesis of the spine without exploration of the canal contents, by thoracotomy |
| **LECA003** | Osteosynthesis of the vertebral column with exploration of the canal contents, by thoracotomy |
| **LECA005** | Osteosynthesis and/or anterior arthrodesis of the spine without exploration of the canal contents, by thoraco-phreno-laparotomy |
| **LECA006** | Osteosynthesis and/or anterior arthrodesis or epiphysiodesis of the spine, by thoracoscopy |
| **LECC001** | Spinal osteosynthesis with exploration of canal contents, by laparotomy or lumbotomy |
| **LFCA004** | Osteosynthesis and/or anterior arthrodesis of the spine without exploration of the canal contents, by laparotomy or lumbotomy |
| **LFCA005** | Osteosynthesis and/or anterior arthrodesis of the spinal column, by laparoscopy or retroperitoneoscopy |
| **LFCC001** | Osteosynthesis of the spinal column with exploration of the canal contents, by anterior or anterolateral cervicotomy and by posterior approach |
| **LDCA012** | Osteosynthesis of the spinal column with exploration of the canal contents, by thoracotomy and posterior approach |
| **LECA002** | Osteosynthesis of the spinal column with exploration of the canal contents, by thoraco-phreno-laparotomy and posterior approach |
| **LECA004** | Osteosynthesis of the spinal column with exploration of the canal contents, by thoraco-phreno-laparotomy and posterior approach |
| **LFCA003** | Osteosynthesis of the spinal column with exploration of the canal contents, by laparotomy or lumbotomy and by posterior approach |
| **LHCA001** | Osteosynthesis of the spine without exploration of the canal contents, using an anterior or posterior approach |
| **LFDA001** | Posterior and/or posterolateral arthrodesis of an unreduced lumbar spondylolisthesis, with root release and osteosynthesis, using a posterior approach |
| **LFDA002** | Intercorporeal arthrodesis of large displacement lumbar spondylolisthesis with reduction, with osteosynthesis, using a posterior translombosacral approach |
| **LFDA003** | Posterior and/or posterolateral arthrodesis of lumbar spondylolisthesis without reduction, with root release, without osteosynthesis, using a posterior approach |
| **LFDA005** | Posterior and/or posterolateral arthrodesis of lumbar spondylolisthesis without reduction, without root release, without osteosynthesis, using a posterior approach |
| **LFDA006** | Posterior and/or posterolateral arthrodesis of lumbar spondylolisthesis with reduction, with osteosynthesis, using a posterior approach |
| **LFDA007** | Posterior and/or posterolateral arthrodesis of lumbar spondylolisthesis without reduction, without root release, with osteosynthesis, using a posterior approach |
| **LFDA008** | Large displacement lumbar spondylolisthesis arthrodesis with reduction, with osteosynthesis, by laparotomy and posterior approach |
| **LFDA009** | Intercorporeal arthrodesis of a lumbar spondylolisthesis with reduction, with osteosynthesis, via a posterior approach |
| **LFDA010** | Arthrodesis of a lumbar spondylolisthesis without reduction, by laparotomy and posterior approach |
| **LFDA011** | Large displacement lumbar spondylolisthesis arthrodesis with reduction, with osteosynthesis, by laparotomy |
| **LFDA013** | Arthrodesis of lumbar spondylolisthesis without reduction, by laparotomy |
| **LFDA014** | Intercorporeal arthrodesis of a large displacement lumbar spondylolisthesis without reduction, with osteosynthesis, using a posterior translombosacral approach |
| **LEMA001** | Instrumental correction of a flexible spinal deformity with arthrodesis of 6 or more vertebrae, by thoraco-phreno-laparotomy |
| **LEMA002** | Instrumental correction of a flexible spinal deformity with arthrodesis of 3 to 5 vertebrae, by thoracotomy |
| **LEMA003** | Instrumental correction of a flexible spinal deformity with arthrodesis of 3 to 5 vertebrae, by thoraco-phreno-laparotomy |
| **LEMA004** | Instrumental correction of a flexible spinal deformity with arthrodesis of 6 or more vertebrae, by thoracotomy |
| **LFMA001** | Instrumental correction of a flexible spinal deformity with arthrodesis of 3 to 5 vertebrae, by lumbotomy |
| **LHMA003** | Instrumental correction of a flexible spinal deformity with arthrodesis of 6 to 9 vertebrae, using a posterior approach |
| **LHMA004** | Instrumental correction of a flexible spinal deformity with arthrodesis of 6 to 9 vertebrae, using a posterolateral approach |
| **LHMA006** | Instrumental correction of a flexible spinal deformity with arthrodesis of 3 to 5 vertebrae, using a posterior approach |
| **LHMA011** | Instrumental correction of a flexible spinal deformity without arthrodesis, using a posterior approach |
| **LHMA013** | Instrumental correction of a flexible spinal deformity with arthrodesis of 6 to 9 vertebrae via a posterior approach, with resection of 3 or more ribs |
| **LHMA014** | Instrumental correction of a flexible spinal deformity with arthrodesis of 10 or more vertebrae via a posterior approach, with resection of 3 or more ribs |
| **LHMA015** | Instrumental correction of a flexible spinal deformity with arthrodesis of 10 or more vertebrae, using a posterior approach |
| **LDPA008** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, by cervicotomy |
| **LDPA009** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, by cervicothoracotomy |
| **LDPA010** | Occipitoaxoid osteotomy or arthrectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, by intraoral approach or cervicotomy |
| **LEPA001** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 1 to 3 vertebrae, by thoraco-phreno-laparotomy |
| **LEPA002** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 1 to 3 vertebrae, by thoracotomy |
| **LEPA003** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, on 1 to 3 vertebrae, by thoracotomy |
| **LEPA004** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, on 4 or more vertebrae, by thoraco-phreno-laparotomy |
| **LEPA005** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 4 or more vertebrae, by thoraco-phreno-laparotomy |
| **LEPA006** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 4 or more vertebrae, by thoracotomy |
| **LEPA007** | Anterior osteotomy or total discectomy for rigid spinal deformity, with arthrodesis and instrumental correction, on 4 or more vertebrae, by thoracotomy |
| **LEPA008** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, on 1 to 3 vertebrae, by thoraco-phreno-laparotomy |
| **LEPA009** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 1 to 3 vertebrae, by laparotomy or lumbotomy |
| **LFPA001** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 1 to 3 vertebrae, by laparotomy or lumbotomy |
| **LFPA002** | Anterior osteotomy or total discectomy for rigid deformity of the spine, with arthrodesis and instrumental correction, on 1 to 3 vertebrae, by laparotomy or lumbotomy |
| **LFPA003** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 4 or more vertebrae, by laparotomy or lumbotomy |
| **LHFA001** | Anterior osteotomy or total discectomy for rigid spinal deformity with arthrodesis, without instrumental correction, on 4 or more vertebrae, by laparotomy or lumbotomy |
| **LHFA003** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis and instrumental correction, on 3 to 5 vertebrae, using a posterior approach |
| **LHFA013** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis and instrumental correction, on 10 or more vertebrae, via a posterior approach, with resection of 3 or more ribs |
| **LHFA025** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis, instrumental correction and transpedicular anterior osteotomy, on 6 to 9 vertebrae, using a posterior approach |
| **LHFA027** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis, instrumental correction and transpedicular anterior osteotomy, on 10 or more vertebrae, using a posterior approach |
| **LHFA028** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis, instrumental correction and transpedicular anterior osteotomy, on 3 to 5 vertebrae, using a posterior approach |
| **LHFA029** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis and instrumental correction, on 10 or more vertebrae, using a posterior approach |
| **LDPA001** | Bilateral total arthrectomy and/or posterior osteotomy for rigid spinal deformity with arthrodesis and instrumental correction, on 6 to 9 vertebrae, using a posterior approach |
| **LDPA002** | Spinal cord decompression for occipitocervical junction deformity, by anterolateral cervicotomy |
| **LDPA003** | Medullary decompression for occipitocervical junction malformation, without dural opening, by posterior approach |
| **LDPA004** | Spinal cord decompression for occipitocervical junction malformation, with osteosynthesis, using a posterior approach |
| **LDPA005** | Medullary decompression for occipitocervical junction deformity, via intraoral approach |
| **LHMA007** | Vertebral laminoplasty without exploration of the intradural contents, using a posterior or posterolateral approach |
| **LHMA016** | Vertebral laminoplasty with exploration of the intradural contents and plasty of the dura mater, using a posterior or posterolateral approach |
| **LHPA003** | Spinal laminotomy without exploration of the intradural contents, using a posterior or posterolateral approach |
| **LHPA006** | Vertebral laminotomy with exploration of the intradural contents and plasty of the dura mater, using a posterior or posterolateral approach |
| **LHPA010** | Vertebral laminotomy with exploration of the intradural contents without plasty of the dura mater, using a posterior approach or a posterolateral approach |
| **LHFA016** | Vertebral laminectomy without exploration of the intradural contents, using a posterior or posterolateral approach |
| **LHFA019** | Vertebral laminectomy with exploration of the intradural contents and plasty of the dura mater, using a posterior or posterolateral approach |
| **LHFA024** | Vertebral laminectomy with exploration of intradural contents without dural plasty, using a posterior or posterolateral approach |
| **LDFA003** | Bilateral total cervical laminarthrectomy, using a posterior approach |
| **LDFA004** | Unilateral total cervical laminarthrectomy with osteosynthesis, using a posterior approach |
| **LDFA005** | Unilateral total cervical lamiarthrectomy without osteosynthesis, using a posterior approach |
| **LFFA001** | Bilateral total lumbar or lumbosacral lamiarthrectomy, using a posterior approach |
| **LFFA005** | Unilateral total lumbar or lumbosacral lamiarthrectomy with osteosynthesis, using a posterior approach |
| **LFFA006** | Unilateral total lumbar or lumbosacral laminarthrectomy without osteosynthesis, using a posterior approach |
| **LDAA001** | Bilateral recalibration of the cervical spine, using a posterior approach |
| **LDAA002** | Unilateral recalibration of the cervical spine, using a posterior approach |
| **LFAA001** | Unilateral recalibration of the lumbar or lumbosacral spine, using a posterior approach |
| **LFAA002** | Bilateral recalibration of the lumbar or lumbosacral spine, using a posterior approach |
| **LDFA002** | Uncectomy [Resection of the uncus] or unilateral foraminotomy of a vertebra, by anterior cervicotomy |
| **LDPA006** | Corporotomy [Somatotomy] of a vertebra for spinal cord decompression, by anterior or anterolateral cervicotomy |
| **LDPA007** | Corporotomy [Somatotomy] of a vertebra for spinal cord decompression, with arthrodesis and/or osteosynthesis, by anterior or anterolateral cervicotomy |
| **LDFA009** | Partial vertebral corporectomy, by anterior or anterolateral cervicotomy |
| **LDFA012** | Total vertebral corporectomy, by anterior or anterolateral cervicotomy |
| **LEFA004** | Total vertebral corporectomy, by thoraco-phreno-laparotomy |
| **LEFA006** | Total vertebral corporectomy, by thoracotomy |
| **LEFA007** | Partial vertebral corporectomy, by thoraco-phreno-laparotomy |
| **LEFA008** | Corporectomy of a malformed vertebra, by thoraco-phreno-laparotomy |
| **LEFA010** | Partial vertebral corporectomy, by thoracotomy |
| **LEFA012** | Corporectomy of a malformed vertebra, by thoracotomy |
| **LEFA014** | Partial or total vertebral corporectomy, by thoracoscopy or by thoracotomy with thoracoscopic preparation |
| **LFFA008** | Corporectomy of a malformed vertebra, by laparotomy or lumbotomy |
| **LFFA009** | Partial vertebral corporectomy, by laparotomy or lumbotomy |
| **LFFA013** | Total vertebral corporectomy, by laparotomy or lumbotomy |
| **LFFA014** | Partial or total vertebral corporectomy, by direct approach with preparation by laparoscopy or retroperitoneoscopy |
| **LHFA031** | Total removal of the vertebral arch, using a posterior approach |
| **LDFA010** | Total spondylectomy, by anterior or anterolateral cervicotomy and posterior approach |
| **LEFA001** | Total thoracic spondylectomy, posterior approach |
| **LEFA005** | Total spondylectomy, by thoraco-phreno-laparotomy and posterior approach |
| **LEFA009** | Total spondylectomy, thoracotomy and posterior approach |
| **LFFA012** | Total spondylectomy, by laparotomy or lumbotomy and posterior approach |
| **LGFA001** | Distal removal of the sacrum [Sacrectomy respecting S1 and S2], using a posterior approach |
| **LGFA002** | Proximal removal of the sacrum [Sacrectomy S1 and/or S2], via anterior or posterior approach |
| **LGFA003** | Distal removal of the sacrum [Sacrectomy respecting S1 and S2], using an anterior or posterior approach |
| **LGFA004** | Total removal of the sacrum [Total sacrectomy], using an anterior or posterior approach |
| **LGFA005** | Removal of the coccyx |
| **LGFA006** | Proximal removal of the sacrum [S1 and/or S2 sacrectomy], via anterior and posterior approaches |
| **LDGA001** | Removal of osteosynthesis material from the atlas and/or axis, by cervicotomy |
| **LDGA002** | Removal of osteosynthesis material from the spinal column, by anterior or anterolateral cervicotomy |
| **LEGA001** | Removal of osteosynthesis material from the spinal column, by thoraco-phreno-laparotomy |
| **LEGA002** | Removal of osteosynthesis material from the spine, by thoracotomy |
| **LFGA001** | Removal of osteosynthesis material from the spine, by laparotomy or lumbotomy |
| **LHGA004** | Removal of osteosynthesis material from the spine over 10 or more vertebrae, using a posterior approach |
| **LHGA006** | Removal of osteosynthesis material from the spinal column over 6 to 9 vertebrae, using a posterior approach |
| **LHGA007** | Removal of osteosynthesis material from the spine over 2 to 5 vertebrae, using a posterior approach |
| **LHFH001** | Transcutaneous vertebral bone resection with CT guidance |
| **LHMA008** | Reconstruction of the interarticular isthmus of a vertebra, using a posterior approach |

### Identification of infective events

#### ICD10-codes

|  |  |
| --- | --- |
| **Code** | **Label** |
| M462 | Vertebral osteomyelitis |
| M463 | Vertebral Osteomyelitis - Cervico-dorsal region |
| M465 | Vertebral Osteomyelitis - Dorso-lumbar region |
| T813 | Disunion of a surgical wound, not otherwise classified |
| T814 | Infection after a diagnostic and therapeutic procedure, not elsewhere classified |
| T845 | Infection and inflammatory reaction due to an internal joint prosthesis |
| T846 | Infection and inflammatory reaction due to an internal fixation device [any location] |
| T847 | Infection and inflammatory reaction due to other prostheses, implants and internal orthopedic grafts |
| Y831 | Surgical procedure with implantation of an internal prosthesis causing abnormal reactions of the patient or subsequent complications, without mention of an accident during the procedure |
| M8618 | Other acute osteomyelitis - Other locations |
| M8688 | Other osteomyelitis - Other locations |
| M961 | Post-laminectomy syndrome, not elsewhere classified |
| M968 | Affections du système ostéo-articulaire et des muscles après un acte à visée diagnostique et thérapeutique |
| M969 | Affection of the osteo-articular system and muscles following a diagnostic or therapeutic procedure, unspecified |

#### Natural language query

|  |
| --- |
| "infection du site operatoire"~5 OR "infection plaie opératoire"~5 OR "ISO" OR "infection plaie chirurgicale"~5 OR "infection plaie post-opératoire"~5 OR "infection du site chirurgical"~3 OR "plaie chirurgicale infectée"~3 OR "infection cicatrice"~3 OR "sepsis au niveau de la cicatrice"~3 OR "infection au niveau de la cicatrice"~3 OR "infection au site de l'opération"~3 OR "lavage cicatrice"~5 OR "evacuation peridural"~5 OR "écoulement purulent"~5 OR pyorrhée~2 OR "écoulement séropurulent"~5 OR "écoulement séreux"~3 OR "désunion d'une plaie opératoire"~3 OR "rupture d'une plaie opératoire"~3 OR "cicatrice avec désunion"~3 OR "déhiscence de la cicatrice"~3 OR "disjonction de la cicatrice opératoire"~3 OR "désunion cicatricielle"~3 OR "déhiscence cicatricielle"~3 OR "disjonction de la cicatrice"~3 OR "déhiscence de la cicatrice opératoire"~3 OR "rupture de la plaie"~3 OR "déhiscence d'une plaie opératoire"~3 OR "rupture de la plaie post-opératoire"~3 OR "drainage chirurgical"~3 OR "incision et drainage"~3 OR "incision et évacuation"~3 |

#### CCAM procedures

|  |  |
| --- | --- |
| **Code** | **Procedure** |
| AFPA001 | Cleaning of postoperative spinal and/or paravertebral infectious lesions, by direct approach |
| LHPA004 | Debridement of an infectious or ossifluent spinal lesion using a posterior approach |
| QZJA009 | Evacuation of superficial collections of skin using a direct approach |
| QZJA011 | Evacuation of deep collections of skin and soft tissue, using a direct approach |
| QZJB002 | Evacuation of superficial and/or deep collections of skin and soft tissue, by transcutaneous route without guidance |
| QZJA001 | Trimming and/or suturing of deep skin and soft tissue wounds over 10 cm long, outside the face and hand |
| LHGA00\* | Removal of osteosynthesis material from the spinal column |

#### Implants

|  |
| --- |
| **Implant label** |
| spinal artificial joint ligament, replacement or reinforcement, implant and |
| osteosynthesis, lockable centromedullary nail, long bones, associated fractures |
| osteosynthesis, non resorbable tendon or ligament anchorage system, tornier |
| osteosynthesis, self-tapping screw, any type |
| osteosynthesis, drilled screw |
|  |
| spine, cervical cage - locking - bone substitute, medicrea |
| spine, cervical cage - locking - bone substitute,medtronic |
| spine, cervical cage + locking - bone substitute,medicrea |
| spine, cervical cage + locking - bone substitute,spineart |
| spine, interbody cage or equivalent,adsm |
| spine, interbody cage or equivalent,eurospine |
| spine, interbody cage or equivalent,global s |
| spine, interbody cage or equivalent,h.p.i. |
| spine, interbody cage or equivalent,medicrea |
| spine, interbody cage or equivalent,medtronic |
| spine, interbody cage or equivalent,osd |
| spine, interbody cage or equivalent,spineart |
| spine, interbody cage or equivalent,spinevision |
| spine, interbody cage or equivalent,stryker |
| spine, interbody cage or equivalent,zimmer biomet |
| spine, thor/lumbar cage - locking - bone substitute,h.p.i. |
| spine, thor/lumbar cage - locking - bone substitute,medtronic |
| spine, thoracic/lumbar cage + locking - bone substitute,spinevision |
| spine, thoracolumbar cage + locking - bone substitute,zimmer biomet |
| spine, anchoring implant, hook, clamp-hook,atf |
| spine, anchorage implant, hook, clamp-hook,medtronic |
| spine, anchorage implant, cerclage wire, cable,medtronic |
| spine, anchorage implant, cervical pedicle screw,medtronic |
| spine, anchoring implant, cervical pedicle screw,spineart |
| spine, anchoring implant, pedicle screw,medicrea |
| spine, anchorage implant, pedicle screw,medtronic |
| spine, anchoring implant, pedicle screw,neo medical |
| spine, anchoring implant, pedicle screw,stryker |
| spine, anchoring implant, single specific screw,medicrea |
| spine, anchorage implant, single specific screw,medtronic |
| spine, anchorage implant, single specific screw,stryker |
| spine, longitudinal union implant, frame,medtronic |
| spine, longitudinal union implant, connector,medtronic |
| spine, longitudinal union implant, plate,medtronic |
| spine, longitudinal union implant, plate,spineart |
| spine, longitudinal union implant, plate,stryker |
| spine, longitudinal union implant, plate,zimmer biomet |
| spine, longitudinal union implant, rod,medtronic |
| spine, longitudinal union implant, stem,neo medical |
| spine, longitudinal union implant, stem,stryker |
| spine, transverse union implant,medtronic |

## Characteristics of the 300 random selected cases

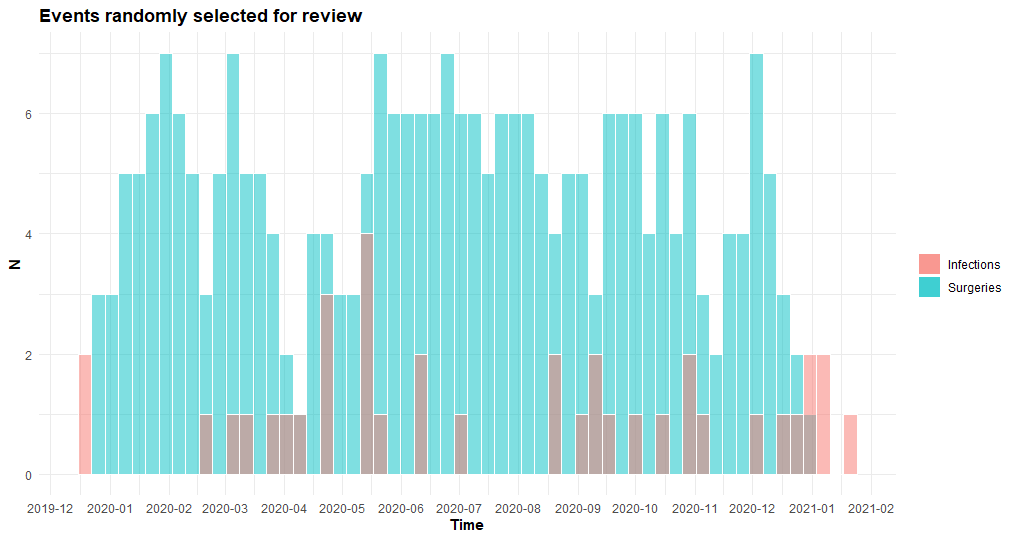


Table 5: primary dispersion measures for the infective events. Randomly selected events does not seem to differ significantly from the cohort’s average.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N | Mean | Sd | Min | Max |
| Randomly selected infections | 38 | 19.6 | 22.2 | 0 | 87 |
| Hospital stays with a notion of infection | 96 | 19.2 | 26.5 | 0 | 111 |

## Metrics Legend

#### Prevalence Metrics

**Apparent Prevalence**: The proportion of cases that tested positive for the condition in the study, regardless of whether they are true or false positives.

**True Prevalence**: The proportion of cases that actually have the condition in the study, confirmed through a gold standard or manual review.

#### Basic Performance Metrics

**Sensitivity**: The proportion of true positives correctly identified by the algorithm. It tells you how good the test is at detecting the condition.

**Specificity**: The proportion of true negatives correctly identified by the algorithm. It tells you how good the test is at avoiding false alarms.

#### Predictive Values

**Positive Predictive Value (PPV):** The proportion of positive test results that are true positives. It answers the question, "If the test says positive, what's the chance it's correct?"

**Negative Predictive Value (NPV):** The proportion of negative test results that are true negatives. It answers the question, "If the test says negative, what's the chance it's correct?"

#### Likelihood Ratios

**Positive Likelihood Ratio**: The ratio of the probability of a positive test result given the presence of the disease to the probability of a positive test result given the absence of the disease.

**Negative Likelihood Ratio**: The ratio of the probability of a negative test result given the presence of the disease to the probability of a negative test result given the absence of the disease.

#### Additional Metrics

**False T+ for True D-:** The proportion of false positives among the true negatives. It tells you how often the algorithm wrongly flags a condition when it's actually not present.

**False T- for True D+:** The proportion of false negatives among the true positives. It tells you how often the algorithm misses a condition when it's actually present.

**False T+ proportion for T+:** The proportion of false positives among all positives (true and false).

**False T- proportion for T-:** The proportion of false negatives among all negatives (true and false).

#### Overall Performance

**Correctly Classified Proportion:** The proportion of all cases (both positive and negative) that the algorithm correctly identified.

**F1**-**score**: The F1 Score is a harmonic mean of precision and recall, offering a balance between the two metrics. It ranges from 0 to 1, where a higher score indicates better performance. An F1 Score closer to 1 indicates a balanced model with good precision and recall. It is particularly useful when class distributions are imbalanced.

**MCC**: The Matthews Correlation Coefficient is a measure of the quality of binary classifications. It takes into account true and false positives and negatives and is a balanced measure even if the classes are of different sizes. The MCC ranges from -1 to 1. An MCC of 1 indicates a perfect prediction, 0 indicates no better than random prediction, and -1 indicates total disagreement between prediction and observation.