Journal Pre-proof

Automated healthcare-associated infection surveillance using an artificial intelligence algorithm

R.P. dos Santos, D. Silva, A. Menezes, S. Lukasewicz, C.H. Dalmora, O. Carvalho, J. Giacomazzi, N. Golin, R. Pozza, T.A. Vaz

Please cite this article as: dos Santos R, Silva D, Menezes A, Lukasewicz S, Dalmora C, Carvalho O, Giacomazzi J, Golin N, Pozza R, Vaz T, Automated healthcare-associated infection surveillance using an artificial intelligence algorithm, *Infection Prevention in Practice*, https://doi.org/10.1016/j.infpip.2021.100167.

Credit Author Statement

dos Santos RP: Conceptualization, methodology, formal analysis, writing original draft, writing review, and editing, visualization and supervision. Silva D: software, formal analysis, and writing original draft. Menezes A: data curation, software, and formal analysis. Lukasewicz S: data curation, validation, formal analysis, writing original draft and visualization. Dalmora CH: validation, visualization. Carvalho O: validation, visualization. Giacomazzi J: Conceptualization, methodology, data curation, writing original draft, visualization, Project administration, funding acquisition. Golin N: data curation, visualization. Pozza R: visualization, funding acquisition. Vaz TA: Conceptualization, methodology, data curation, software, formal analysis, resources, writing original draft, writing review, and editing.

Automated healthcare-associated infection surveillance using an artificial intelligence algorithm

dos Santos RP[1], Silva D[1], Menezes A[1], Lukasewicz S[1], Dalmora CH[1], Carvalho O[1], Giacomazzi J[2], Golin N[2], Pozza R[2], Vaz TA[1].

[1] Qualis Soluções em Infectologia; [2] Hospital Tacchini.

Running title: Automated Infection Surveillance

Corresponding Author:

Rodrigo Pires dos Santos.

Oswaldo Aranha, 1022. Room 1101.

Porto Alegre, RS. 90035-191

Brazil.

e-mail: rodrigo@portalqualis.com.br

Summary. Healthcare-associated infections (HAIs) are among the most common adverse events in hospitals. We used artificial intelligence (AI) algorithms for infection surveillance in a cohort study. The model correctly detected 67 out of 73 patients with HAIs. The final model used a multilayer perceptron neural network achieving an area under receiver operating curve (AUROC) of 90.27%; specificity of 78.86%; sensitivity of 88.57%. Respiratory infections had the best results (AUROC ≥93.47%). The AI algorithm could identify most HAIs. AI is a feasible method for HAI surveillance, has the potential to save time, promote accurate hospital-wide surveillance, and improve infection prevention performance.

Keywords: Infection surveillance; Artificial intelligence; Healthcare-associated infection.

Introduction

At least 1 in every 31 hospitalised inpatients on any given day has a healthcare-associated infection (HAI) [1]. The Centers for Disease Control and Prevention (CDC) recommends HAI surveillance for healthcare facilities. [1] This surveillance should be conducted by an infection control professional (ICP) in an active, patient-based, prospective, priority-directed manner [1].

With recent advances in technology, automated surveillance can improve performance, increase accuracy and contribute to patient safety [2]. A variety of systems are available for healthcare use, including simple rule or complex and machine learning (ML) based algorithms [3,4]. In the near future, the use of these technologies may include identification of patients at risk and promote safer care by personalised infection prevention [5].

Patients and Methods

We conducted a retrospective cohort study for HAI surveillance using an artificial intelligence (AI) algorithm. Hospital Tacchini is a general institution that serves clinical and surgical patients in Southern Brazil. It has 289-beds, including intensive care units for clinical, paediatric and neonatal patients.

The dataset included 18 months of information from the hospital electronic health record (EHR) from all adult (≥ 18 years) inpatients. The data included laboratory results, vital signs, drug prescriptions, imaging, and healthcare worker notes for the duration of the inpatient stay. The dataset was separated into two samples: first, the supervised training samples containing 12 months of data (from January 1st 2017, to December 31st 2017). In this dataset sample the HAI cases had been previously labelled by ICPs. Second, the validation sample dataset containing 6 months of data (from January 1st 2018, to June 3rd 2018). This discriminative source of data means the classification of new data on the basis of similar data in the training set. The outcome for the prediction model was the diagnosis of HAI.

The following infections were included in the study: pneumonia (PNM), ventilator-associated pneumonia (VAP), urinary tract infection (UTI), surgical site infection (SSI), primary bloodstream infection (BSI) and tracheobronchitis (TRACH).

The hospital has a designated infection control teamresponsible for the surveillance of HAI. Two infection control nurses performed the review of HAIs based on positive bacteriological cultures. All patients with positive culture results had their records evaluated for the presence of signs and symptoms of infection. The diagnosis of HAI was based on published criteria from the Centers for the Disease Control and Prevention (CDC) [1]. The gold standard for algorithm training was the diagnosis made by the infection control team, which was reviewed by an independent senior nurse specialised in infection control.

The training of the algorithms was performed using the Python programming language **(Python 3.6 version - www.python.org. Accessed on August 12th, 2017)** using the open-source softwares TensorFlow **(Tensorflow 1.3 version - www.tensorflow.org. Accessed on July 31st, 2019)** and scikit-learn **(sklearn 0.21 version - scikit-learn.org. Accessed on August 14th, 2019)** libraries. For each of the infections, three different approaches were taken: categorical/numerical (vital signs, laboratory results) data search, and natural language processing (NLP) techniques using word embeddings (vector representation of words that are expected to be similar in meaning for text analysis) and bag-ofwords (where a sentence is represented in a multi-set of words). Data preparation was projected to allow each different Artificial Intelligence model to be built using a retrospective dataset of clinical data grouped by segregated intervals of time for each patient. The results of each of the three approaches were added to a dataset containing all the features used to classify the infections individually. Data management and the application of information retrieval and NLP techniques were performed on the Dataiku collaborative data science software platform, version 4.0.5 **(www.dataiku.com. Accessed on August 12nd, 2017)**. The Dataiku platform enabled data extraction, transformation and loading (ETL) processes and free text information retrieval activities through the Python and SQL programming language. The following algorithms available in the scikit-learn and TensorFlow libraries were used: multinomial naive Bayes, random forest, multilayer perceptron (MLP) and separable convolutional neural network (SepCNN). These were included as part of the analysis because they have parameters for weight and cost adjustment, needed to handle unbalanced datasets and achieve the best performance.

Each criterion was predicted using different models, comparing validation results for three different classification approaches: random forest (RF), logistic regression, and convolutional neural networks (CNN). The positive predictive value (PPV), negative predictive value (NPV), specificity, sensitivity and accuracy were calculated.

To determine the best decision threshold, the area under the receiver operating curve (AUROC) was used. The AUROC analysis provides tools for diagnosis of two classes of outcomes (infected or not infected). The best prediction model would be represented by a point in the upper left corner of the graph representation (100% specificity and 100% sensitivity).

Approval for this study was granted by Associação Dr. Bartholomeu Tacchini/Hospital Tacchini/RS. The project is registered in the Plataforma Brasil, CAAE number 89737218.0.0000.5305.

Results

The final dataset included 5,105 patients. The dataset consisted of 122,261 patient-days of information (a database of 5 terabytes); 84,655 (69.2%) patient-days were used for training and testing and 37,606 (30.8%) patient-days were for validation. Of these, 117,896 (96.4%) patient-days from patients without HAIs, and 4,365 (3.6%) patient-days from individuals with at least one HAI.

Overall 90 infections were identified from 73 patients (nine patients had more than one infection): 29 BSI (32.2%); 25 UTI (27.8%); 12 PNM (13.3%); 9 VAP (10.0%); 8 SSI (88.9%); 7 TRACH (7.8%). The

model correctly detected 67 out of the 73 patients with HAIs. Of the 5,032 patients who had no infection, 4,637 patients were correctly classified as non-infected, and 395 patients represented false-positive results.

The best infection classification occurred through the MLP neural network, reaching an AUROC of 90.27% (SD = 0.15), a PPV of 8.92% (SD = 0.07), a specificity of 78.86% (SD = 0.17), an NPV of 99.66% (SD = 0.01), a sensitivity of 88.57% (SD = 0.51), and an accuracy of 79.08% (SD = 0.16) for the detection of at least one infection (Table I).

In general, the best results for the AUROC were detected by analysis of the categorical/numeric (structured) data for VAP (AUROC, 93.47%; SD, 0.13%), pneumonia (AUROC 95.67%; SD - 0.15%), UTI (AUROC, 88.87%; SD, 0.33%), BSI (AUROC, 88.28%; SD, 0.19%) and SSI (AUROC, 85.74%; SD, 0.52%). For tracheobronchitis, the best result was the unstructured natural language text analysis (AUROC, 98.62%; SD, 0.06%, Figure 1).

Discussion

The time required for the surveillance of HAI by ICPs can be considerable. One study evaluating different HAI surveillance found that the time required for data collection for the review health records was 18 hours per 100 beds per week. This study also reported that the review of microbiology reports by ICPs with the clinical teams identified the highest proportion of HAIs (71%) [6]. A survey in Australia, showed that ICPs spent 36% of their time undertaking surveillance. [7]. The number of beds in the healthcare provider influenced time on surveillance . One study showed variable results from 33% to 64% of hospitals using hospital-wide surveillance of BSI, PNM, UTI, SSI, from 297 hospitals in Europe [8]. Laboratory-based surveillance used in our hospital requires less time, but it has lower sensitivity and is also dependent on specimen collection, laboratory expertise and resources. An automated sensitive method based on CDC criteria could be an acceptable approach for hospital-wide surveillance in terms of the amount of time saved and accuracy, especially in hospitals with limited resources.

Until recently, much of the work in HAI surveillance has been dependent on manual work from a trained ICP and surveillance has been targeted as it is not possible to incorporate all the data from all possible sources. There is a range of electronic surveillance methods available for HAIs. A systematic review found excellent results for electronic surveillance in different hospital areas from different countries; the sensitivity varied from 60-100%; the specificity ranged from 37-100%; the PPV varied from 7.4-100%; and the NPV varied from 85-100% [3]. A recent systematic review suggested that machine learning-based models had relatively high performance for specificity and negative predictive value for HAI surveillance [4]. Our AI model was designed to be a sensitive primary source of information for ICPs in HAI surveillance. Cohen et al., have reached similar results, 92% sensitivity, 72% specificity, 74% of accuracy and have proposed methods to increase sensitivity in unbalanced data [9]. In our study, the parameter CW (Class_Weight) was used to minimise the bias between clean and infected cases in an unbalanced dataset. For instance, a CW = 1/0.01, implies in assigning a weight of 1 to a positive (infected) classification and 0.01 to a negative (clean) classification.

Surveillance definitions are different from the clinical diagnosis made by the attending physician. Data from clinical notes, demographic data, structured information from the laboratory, chest X-rays, and vital signs are essential for surveillance and include all 33 criteria defined by the CDC-NHSN [2,6]. Our algorithm incorporated structured (vital signs, laboratory results, drug prescription) and unstructured data (healthcare workers notes, imaging results) clinical sources of data to enhance performance, as clinical data appear to be more sensitive and consistent than laboratory-based surveillance only, or administrative data [10].

Considering the recommended ICP ratio, the complexity of healthcare and the key role of ICPs in delivering high-value care, AI algorithms have the potential to save time for HAI diagnosis, promote accurate hospital-wide surveillance, and improve infection prevention performance. This automated model could output the preselected criteria identified in the patient records to the ICP's final manual ascertainment, increasing the predictive positive value and accuracy. This semi-automated process (machine and manual) also serves as the algorithm stewardship done by ICPs, which includes additional layers of quality control in the whole chain of HAI surveillance.

Our study has a number of limitations. This is a single centre study, from a single database that represents a unique cohort of adult inpatients. Furthermore, the EHR data and the way they are transcribed and documented will vary between healthcare providers and may impact the generalisation of our results. On the other hand, the use of machine learning processes can prevent the need for generic rules in the surveillance process, which could facilitate implementation and generalisation for different scenarios, such as cardiac surgery or cancer centres. The gold standard for the HAI identification was the diagnosis determined by the ICPs. Human error can occur independent of the method used. The ICPs used laboratory culture results as the basis for HAI surveillance. Laboratory-based surveillance may still miss one third of infections [3]. We were not able to review most of the false-positive results. This may have identified the estimated third of missed infections by manual review, which would in turn influence the PPV in favour of the AI algorithm.

In conclusion, the AI algorithm identified most cases of HAIs with acceptable sensitivity and accuracy, reaching high negative predictive values. The access and implementation of these AI technologies are still a challenge. Future studies comparing manual and automated AI based models are needed to investigate potential time and cost savings and improved efficiency.

Conflicts of interest

RP Santos, C Dalmora and Vaz Tiago are the owners of Qualis startup.
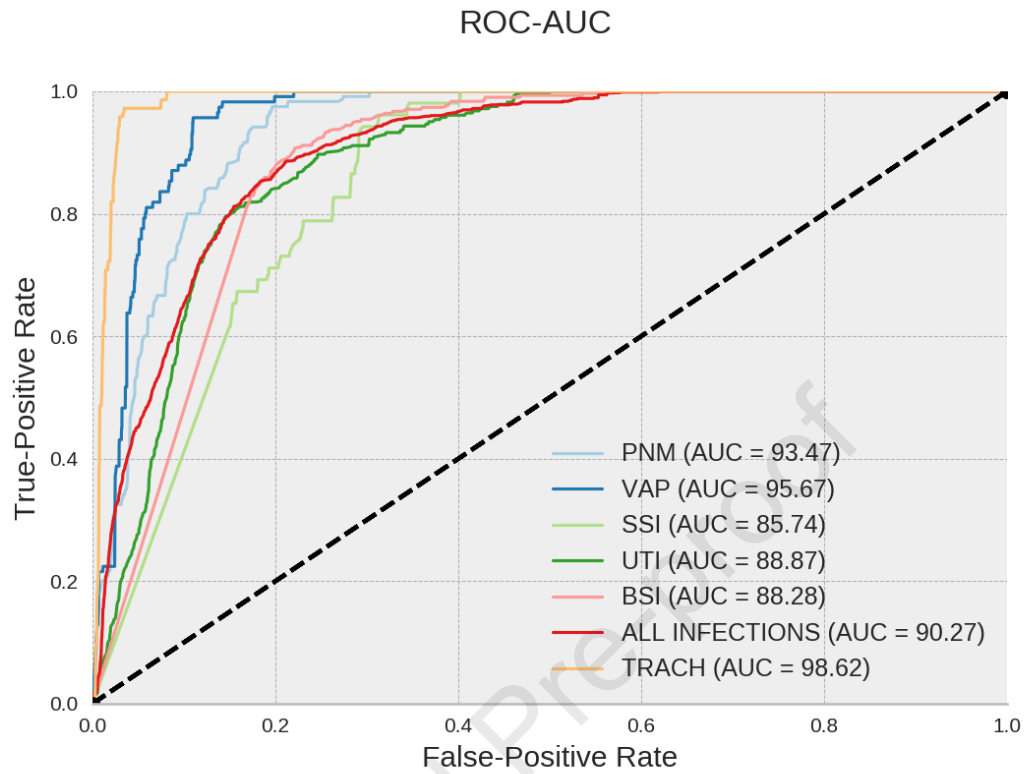
Funding

References

[1]   Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control 2008;36:309e332.

[2]   Woeltje KF, Lin MY, Klompas M, Wright MO, Zuccotti G, Trick WE. Data requirements for electronic surveillance of healthcare-associated infections. Infect Control Hosp Epidemiol 2014;35:1083-91.

[3]   Freeman R, Moore LS, García Álvarez L, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. J Hosp Infect 2013;84:106-19.

[4]   Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature. J Infect Public Health 2020;13:1061-77.

[5]   Gastmeier P. From 'one size fits all' to personalized infection prevention. J Hosp Infect 2020;104:256-60.

[6]   Glenister HM, Taylor LJ, Bartlett CL, Cooke EM, Sedgwick JA, Mackintosh CA. An evaluation of surveillance methods for detecting infections in hospital inpatients. J Hosp Infect 1993;23:229-42

[7]   Mitchell BG, Hall L, Halton K, MacBeth D, Gardner A. Time spent by infection control professionals undertaking healthcare associated infection surveillance: a multicentred cross sectional study. Infect Dis Health 2016; 21:36–40.

[8]   Hansen S, Schwab F, Zingg W, Gastmeier P; The Prohibit Study Group. Process and outcome indicators for infection control and prevention in European acute care hospitals in 2011 to 2012 - Results of the PROHIBIT study. Euro Surveill 2018; 23.

[9]   Cohen G, Sax H, Geissbuhler A. Novelty detection using one-class Parzen density estimator. An application to surveillance of nosocomial infections. Stud Health Technol Inform 2008;136:21-6

[10]   Rhee C, Kadri S, Huang SS, Murphy MV, Li L, Platt R,, et al. Objective sepsis surveillance using electronic clinical data. Infect Control Hosp Epidemiol 2016;37:163– 171.

Table I Final results based on the algorithm used for each HAI

	Algorithm	Dataset	AUROC (SD)	NPV (SD)	PPV (SD)	Sensitivity (SD)	Specificity (SD)	Accuracy (SD)
Global	MLP	Categorical/ Numerical	90.27% (0.15)	99.66% (0.01)	8.92% (0.07)	88.57% (0.51)	78.86% (0.17)	79.0% (0.16)
Pneumonia	RF	Categorical/ Numerical	93.47% (0.13)	99.99% (0.0)	1.59% (0.05)	97.31% (0.77)	80.66% (0.65)	80.7% (0.64)
Ventilator-associated Pneumonia	RF	Categorical/ Numerical	95.67% (0.15)	99.99% (0.0)	2.62% (0.02)	95.69% (0.48)	89.02% (0.05)	89.0% (0.05)
Surgical Site Infection	MNB	Categorical/ Numerical	85.74% (0.52)	99.99% (0.0)	0.44% (0.01)	95.09% (1.35)	70.03% (1.06)	70.0% (1.06)
Urinary Tract Infection	RF	Categorical/ Numerical	88.87% (0.33)	99.85% (0.03)	3.42% (0.49)	84.0% (4.3)	81.49% (4.04)	81.5% (3.98)
Bloodstream Infection	MNB	Categorical/ Numerical	88.28% (0.19)	99.9% (0.01)	3.26% (0.03)	90.85% (0.67)	77.91% (0.12)	78.0% (0.12)
Tracheobronch-itis	RF	Text Bag-of-Words (TF-IDF)	98.62% (0.06)	99.99% (0.0)	5.26% (0.30)	97.22% (0.93)	96.63% (0.18)	96.6% (0.18)

MLP – Multilayer Perceptron; RF – Random Forest; MNB - Multinomial Naive Bayes; AUROC - Area Under the Receiver Operating Characteristic Curve; NPV – Negative Predictive Value; PPV- Positive Predictive Value; SD – Standard Deviation. TF-IDF – Term Frequency-Inverse Document Frequency.

Figure 1 AUROC for detecting each infection by Multilayer Perceptron Neural Network

ROC-AUC

PNM (AUC = 93.47)
VAP (AUC = 95.67)
SSI (AUC = 85.74)
UTI (AUC = 88.87)
BSI (AUC = 88.28)
ALL INFECTIONS (AUC = 90.27)
TRACH (AUC = 98.62)

PNM – pneumonia; VAP - ventilator-associated pneumonia; UTI - urinary tract infection; SSI - surgical site infection; BSI - bloodstream infection; TRACH – tracheobronchitis; ROC - receiver operating curve; AUC – area under curve.