Machine Learning for Detection of Correct Peripherally Inserted Central Catheter Tip Position from Radiology Reports in Infants

Manan Shah[1] Derek Shu[2] V. B. Surya Prasath[3] Yizhao Ni[3,4] Andrew H. Schapiro[5,6] Kevin R. Dufendach[1,2,3]

[1] Division of Neonatal and Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States
[2] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States
[3] Division of Bioinformatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States
[4] Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States

Appl Clin Inform 2021;12:856–863.

Address for correspondence Manan Shah, MD, Division of Neonatal and Pulmonary Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue MLC 7009, Cincinnati, OH 45229, United States (e-mail: mas476@gmail.com).

[5] Department of Radiology and Medical Imaging, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States
[6] Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States

Abstract

Background In critically ill infants, the position of a peripherally inserted central catheter (PICC) must be confirmed frequently, as the tip may move from its original position and run the risk of hyperosmolar vascular damage or extravasation into surrounding spaces. Automated detection of PICC tip position holds great promise for alerting bedside clinicians to noncentral PICCs.

Objectives This research seeks to use natural language processing (NLP) and supervised machine learning (ML) techniques to predict PICC tip position based primarily on text analysis of radiograph reports from infants with an upper extremity PICC.

Methods Radiographs, containing a PICC line in infants under 6 months of age, were manually classified into 12 anatomical locations based on the radiologist's textual report of the PICC line's tip. After categorization, we performed a 70/30 train/test split and benchmarked the performance of seven different (neural network, support vector machine, the naïve Bayes, decision tree, random forest, AdaBoost, and K-nearest neighbors) supervised ML algorithms. After optimization, we calculated accuracy, precision, and recall of each algorithm's ability to correctly categorize the stated location of the PICC tip.

Results A total of 17,337 radiographs met criteria for inclusion and were labeled manually. Interrater agreement was 99.1%. Support vector machines and neural networks yielded accuracies as high as 98% in identifying PICC tips in central versus noncentral position (binary outcome) and accuracies as high as 95% when attempting to categorize the individual anatomical location (12-category outcome).

Conclusion Our study shows that ML classifiers can automatically extract the anatomical location of PICC tips from radiology reports. Two ML classifiers, support vector machine (SVM) and a neural network, obtained top accuracies in both binary and multiple category predictions. Implementing these algorithms in a neonatal intensive care unit as a clinical decision support system may help clinicians address PICC line position.

Keywords
- supervised machine learning
- natural language processing
- medical error reduction
- radiology information systems
- patient safety

received
February 14, 2021
accepted after revision
July 25, 2021

DOI https://doi.org/10.1055/s-0041-1735178.
ISSN 1869-0327.

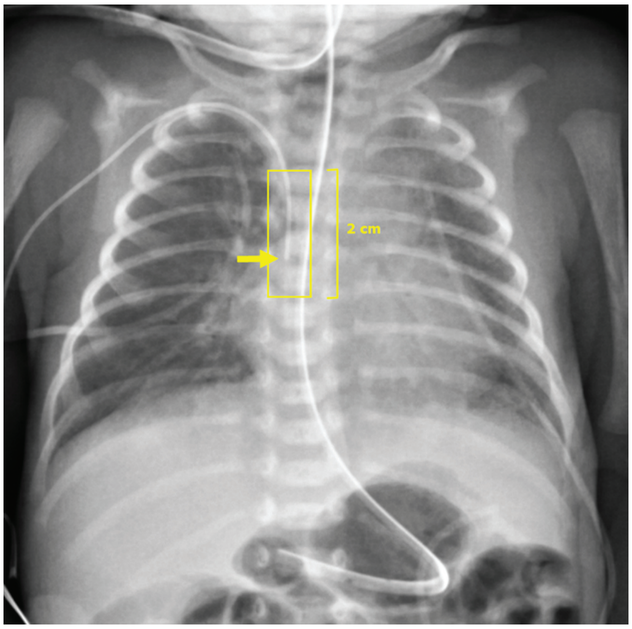

Fig. 1 An approximately 1 kg baby with a PICC. The yellow arrow points to the tip of the upper extremity PICC in an appropriate position in the SVC. The yellow boxed area represents the location where the tip position would be considered appropriate (SVC and the SVC/RA junction). PICC, peripherally inserted central catheter; RA, right atrium; SVC, superior vena cava.

Background and Significance

Infants in intensive care units (ICUs) frequently require peripherally inserted central catheters (PICC) to provide long-term central venous access for medication, parenteral nutrition, and fluids.[1] Because the PICC tip can unintentionally move from its original position, its position must be confirmed radiographically.[2] An upper extremity PICC line tip is optimally positioned in the superior vena cava (SVC) or at the SVC/right atrium (SVC/RA) junction.[3] This target area can be quite small in neonates, as shown in ►Fig. 1, an infant weighing 1 kg has a target range of only 2 cm where the PICC line would be considered "optimal." In infants smaller than 1 kg, this target location could be substantially smaller.

PICC tips outside of this location run the risk of vascular damage due to fluid composition, extravasation into surrounding spaces, precipitation of arrhythmias, and other life-threatening complications.[4–6] As the optimum area of placement is small, even small movements or changes in position may cause significant displacement. Given the PICC may be necessary for weeks to months, there is ample opportunity for displacement.

In the neonatal ICU (NICU), many X-rays are obtained for indications other than confirming central line placement (e.g., to evaluate pneumonia, effusion, pneumothorax, and others), and the PICC position is therefore incidental to the main purpose. Studies have shown that incidental findings are often missed or have delayed follow-up,[7] and there have been creative solutions (including some using natural language processing [NLP]) to help alleviate this problem.[8] A system that can automatically identify when a PICC tip has moved and alert the clinician to this change may help decrease instances of prolonged PICC tip malpositioning and reduce risk of patient harm.

Regardless of the indication for the radiograph, pediatric radiologists at our institution identify the location of any lines and support devices, including PICC lines, in the radiograph textual report. This report is subsequently evaluated by the clinician to assess the adequacy of PICC tip position. The attending neonatologist ultimately documents in the daily progress note whether the PICC tip is appropriately positioned.

NLP is a set of algorithms that allow computers to extract data from unstructured text.[9–11] The algorithms systematically split free text into individual segments (words and phrases), correct spelling mistakes, standardize abbreviations, and analyze the syntax and semantics of each individual segment to decide whether a concept or sentiment is present within the free text.[12,13] It is a powerful tool when there is a large set of unstructured data, and it allows us gain unique insights that may be difficult to appreciate otherwise, for example, analyzing patient feedback that is written in natural language.[13] A recent review,[9] examined 67 articles on NLP for radiology, reports that clinical support service was a major category of use. NLP has also been used with progress notes to identify disease processes or to aid in billing and coding.[14,15]

NLP has been demonstrated superior to diagnosis codes in identification of clinical findings such as long bone fractures from radiology reports.[16] It has also been found superior to diagnosis codes in identification of patients treated with pneumonia based on analysis of progress notes.[8] It can also provide quantitative information, such as extracting organ measurements from radiology reports.[17] Recent advancements in open-source NLP frameworks have facilitated rapid deployment of NLP solutions, and the technology is now becoming increasingly accessible for researchers without prior experience.[18]

Machine learning (ML) is the discipline that "focuses on how computers learn from data."[19] ML-based techniques can be supervised or unsupervised.[20] Supervised ML starts with a known output, and the computer attempts to devise an algorithm to match that output. For example, a set of magnetic resonance imaging (MRI) scans with known lesions are presented to the computer, and then the computer is given a new MRI image and asked whether the patient has a lesion. Radiology has been a fruitful avenue in ML, with excellent results with NLP and image analysis.[21,22] In other fields, for example, dermatology, some ML models have achieved similar accuracies as physicians in diagnosing specific conditions.[23] In an ML context, a model is essentially an algorithm that takes a set of inputs and produces an output,[24] and there are numerous models, each with their advantages and disadvantages depending on the type of inputs and outputs that are required (please see ►Supplementary Primer [available in the online version] for a brief overview of ML).

There has been significant advancement in digital health and the use of electronic health records (EHRs), and with that, an increased interest in using automation to assist in improving patient safety[25] and to improve physician efficiency. Artificial intelligence (AI) techniques, especially ML, have been used to detect PICC lines in adults through direct image processing.[26] This study seeks to use NLP and supervised ML to extract upper extremity PICC tip location from radiograph reports from infants. While identifying "central" versus "non-central" location is of primary importance, there is added clinical utility in a 12-category classification, as it may alert clinicians if a tip has moved significantly. For example, a noncentral PICC in the brachiocephalic may be acceptable in certain situations (e.g., if it is used for antibiotic administration), but a change in position to the "neck" from "brachiocephalic" would warrant repositioning.

The Institutional Review Board (IRB) of Cincinnati Children's Hospital Medical Center (CCHMC) deemed this study exempt. To the best of our knowledge, in spite of the existence of many studies in the current literature related to the use of NLP and ML in radiology reports, there have been no prior studies that use NLP and ML to categorize and track PICC tip position in neonates.[7,16,27]

Methods

We obtained deidentified radiology reports from CCH, a quaternary children's hospital, as well as an associated delivery level-III NICU at University of Cincinnati (UC) between September 2015 and August 2019. The radiographs from both institutions are interpreted by the same group of pediatric radiologists which helped with standardization. Inclusion criteria were age less than 6 months at the time of the radiograph and a single, upper extremity PICC. Reports were identified via a radiology database search engine (Illuminate Insight by Softek Illuminate Inc., Overland Park, Kansas, United States), with search criteria including the date range specified above and the presence of the word "PICC" in any part of the report. Chest radiograph reports were excluded if there was no upper extremity PICC present, if a second PICC or other upper extremity, cervical approach for non-PICC central vascular catheter (e.g., extracorporeal membrane oxygenation [ECMO] cannulae and hemodialysis catheter) was identified, or if the patient was outside the age criteria.

We created a web-based tool that displayed each radiology report, a dropdown box to identify the PICC tip location, and an option to exclude the report if it did not meet criteria. Using this tool, two researchers (M.S. and D.S.) manually reviewed each textual report and classified the PICC tip position as 1 of the 12 anatomical locations: cephalic, axillary, subclavian, brachiocephalic, subclavian/brachiocephalic, brachiocephalic/SVC, SVC, SVC/RA, RA, neck, chest, and other. Appropriate position in the SVC and SVC/RA was defined as "central" for the purposes of binary classification. One rater (M.S.) was a neonatal fellow with substantial experience reading radiograph reports, and the other (D.S.) was a medical student with limited experience reading radiograph reports. In instances where there was disagreement between the two raters, a board-certified pediatric radiologist provided the tiebreaker.

To predict PICC position outcomes, we chose seven ML classifiers including the naïve Bayes, support vector machine (SVM, also known as support vector classifier), decision tree, random forest, AdaBoost, K-nearest neighbors (KNN), and a custom-designed neural network. We used Python version 3.7[28] and the Scikit-Learn[29] library for all preprocessing and classification. For each report, we parsed out the section containing the word PICC and converted the text into n-grams (varied based on the algorithm; e.g., bigram for neural network, range 1–5 for SVM) using Scikit-Learn CountVectorizer, followed by scaling the frequency via TfidfTransformer (TF-IDF). We did not remove stop words (e.g., "the" and "not") before n-gram extraction to preserve contextual information such as negation. We trained and tested these ML models for each of the two types of output—first, to predict whether the PICC tip was in a central position (SVC or SVC/RA) or noncentral (binary classification) and, second, to identify the exact location (12-category classification using a multinomial model).

To help limit bias due to disproportionate samples from anatomical locations (unbalanced data) in ML models, we performed a 70/30 train/test split of data in each of the 12 categories. All ML algorithms were trained and optimized on 70% of the data, while 30% of the data were set aside and used for testing. The hyperparameters of the ML classifiers (e.g., numbers of hidden units in neural networks) were optimized on the training dataset with 10-fold cross validation. Finally, the ML models with optimal hyperparameters were trained on the whole training set and evaluated on the test set for validation. We adopted accuracy, precision, and recall evaluation metrics to assess model performance. Out of the three, we chose accuracy as the primary metric to assess model performance (sample code available at: *https://github.com/dufendach-lab/picc-ml/*).

Results

A total of 26,963 reports were assessed for eligibility, and 17,337 radiograph reports met inclusion criteria. ►**Fig. 2** shows the division of radiological reports into training and testing data and corresponding binary and 12-category position. The demographic data are described in ►**Table 1**. Our patients had a slight male preponderance and a significant majority was observed under 2 months of age.

The interrater reliability for determining anatomical location was high, with a Cohen's Kappa score of 0.98 between the two individuals. In post hoc analysis of the 161 (0.91%) reports with disparate classification, the tie-breaking pediatric radiologist agreed with one of the two raters 90 (55.9%) times, with most reasons for difference in classification attributed to unclear (60 instances, or 66.7%) or inconsistent (20 instances, or 22.2%) wording (►**Fig. 3**). In the other 71 (44.1%) cases with interrater disagreement, the radiologist classified the report into a category not chosen by either rater, with 58 (81.7%) of these errors due to complex anatomical classification (e.g., unique cardiac defects that affected the location of the tip). For reference, some examples of these reports are provided in ►**Supplementary Table S1** (available in the online version).

We tested the supervised ML classifiers on the curated radiological reports to see if PICC tip position could be

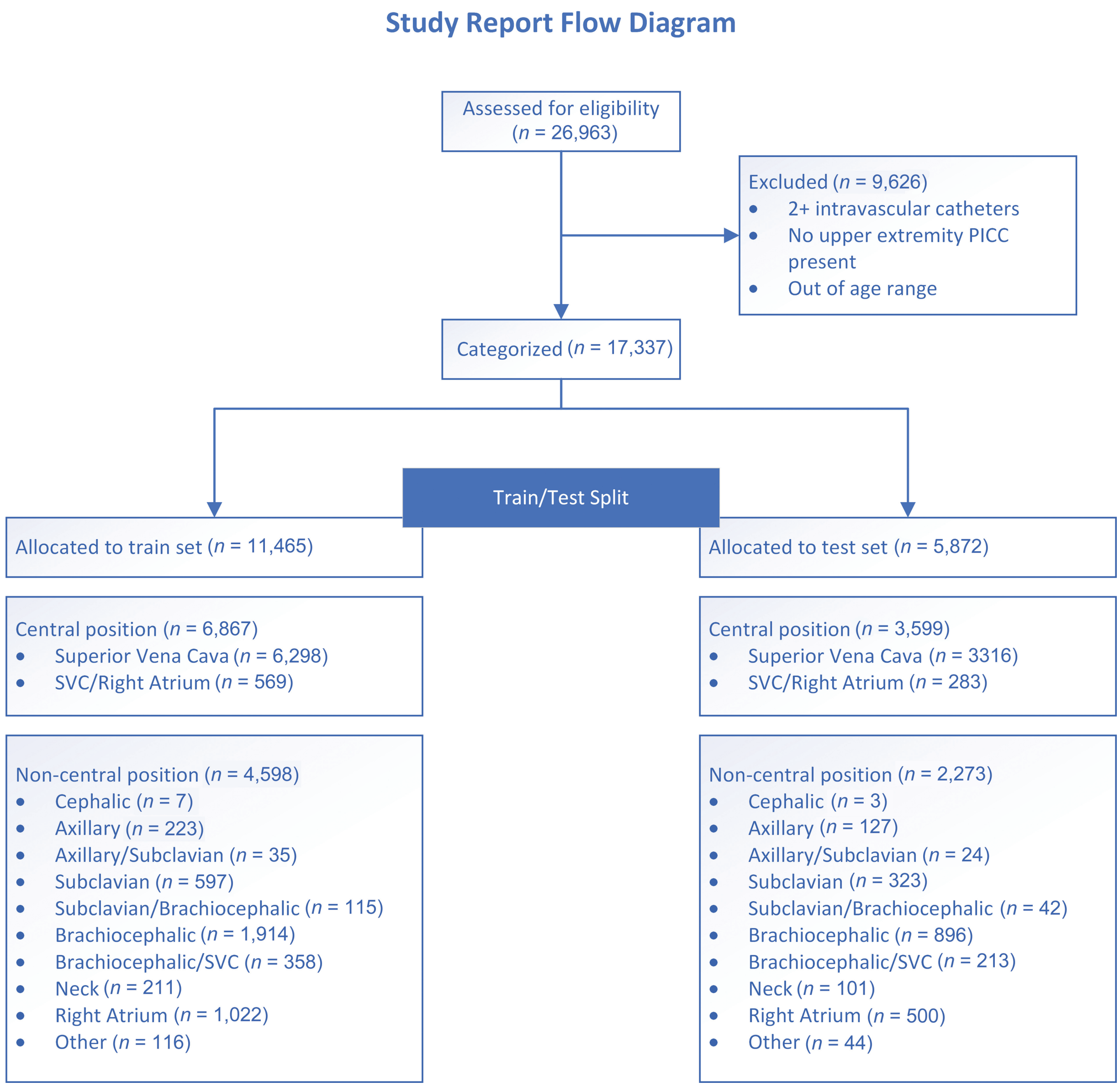

Fig. 2 Number of radiological reports and their corresponding binary outcome considered for the ML classification models. ML, machine learning.

Table 1 Number of radiological reports by age and gender

Demographic	*n* (%)
Gender	
Male	9,451 (54.5)
Female	7,883 (45.4)
Unknown	3 (0.02)
Age (mo)	
0–1	8,202 (47.3)
1–2	2,988 (17.2)
2–3	2,002 (11.5)
3–4	1,763 (10.2)
4–5	1,460 (8.4)
5–6	922 (5.3)

identified accurately. ►**Table 2** presents the benchmarking results for the seven classifiers, with results as high as 98% for binary accuracy and 95% for 12-category accuracy. We showed that multiple models can yield excellent accuracies when analyzing radiology text reports. In our results, the precision and recall were also similar to the accuracy in the binary classification.

These ML algorithms each have their own sets of features and feature importance based on their underlying mechanisms and training data. As a demonstrative example of ML-based classification with respect to salient features, ►**Fig. 4** shows the top 40 features (words or phrases) in the top performing SVM algorithm. The feature importance score for each feature is a value in the interval [−1, 1], with the most positive weights (blue) given for making a central prediction and negative weights (red) for a noncentral prediction. Features such as "SVC" or "projects over the SVC" were strong

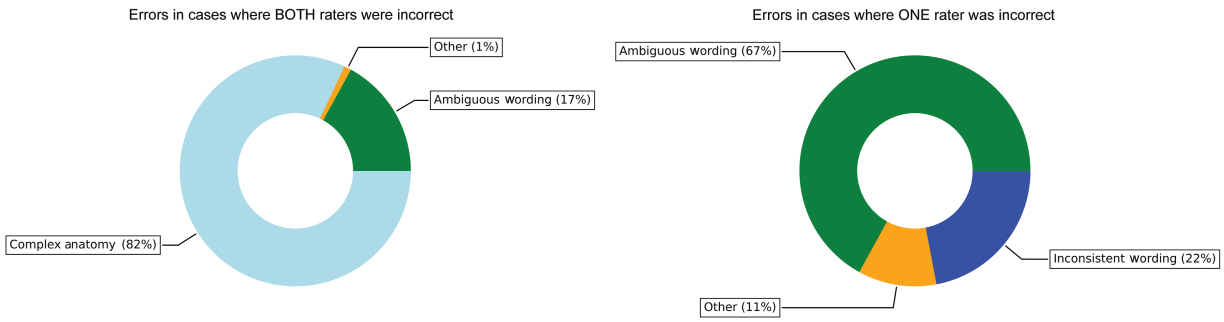

Fig. 3 Analysis of cases ($n = 161$) with interrater disagreement on classification

Table 2 Performance metrics of ML algorithms in binary classification (central vs. noncentral position) and accuracy of specific 12-category position classification

	Binary position (central vs. noncentral)			12-category position
Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
SVM	98 ± 0.36	Central: 98 Noncentral: 98	Central: 97 Noncentral: 99	95 ± 0.56
Neural network	98 ± 0.36	Central: 97 Noncentral: 98	Central: 99 Noncentral: 96	95 ± 0.56
Random forest	97 ± 0.44	Central: 96 Noncentral: 97	Central: 98 Noncentral: 94	90 ± 0.77
The naïve Bayes	94 ± 0.61	Central: 94 Noncentral: 96	Central: 97 Noncentral: 91	84 ± 0.94
Decision tree	93 ± 0.65	Central: 93 Noncentral: 92	Central: 91 Noncentral: 94	85 ± 0.91
AdaBoost	96 ± 0.50	Central: 96 Noncentral: 97	Central: 98 Noncentral: 94	75 ± 1.10
K-nearest neighbor	91 ± 0.73	Central: 91 Noncentral: 89	Central: 95 Noncentral: 86	70 ± 1.17

Abbreviations: ML, machine learning; SVM, support vector machine.

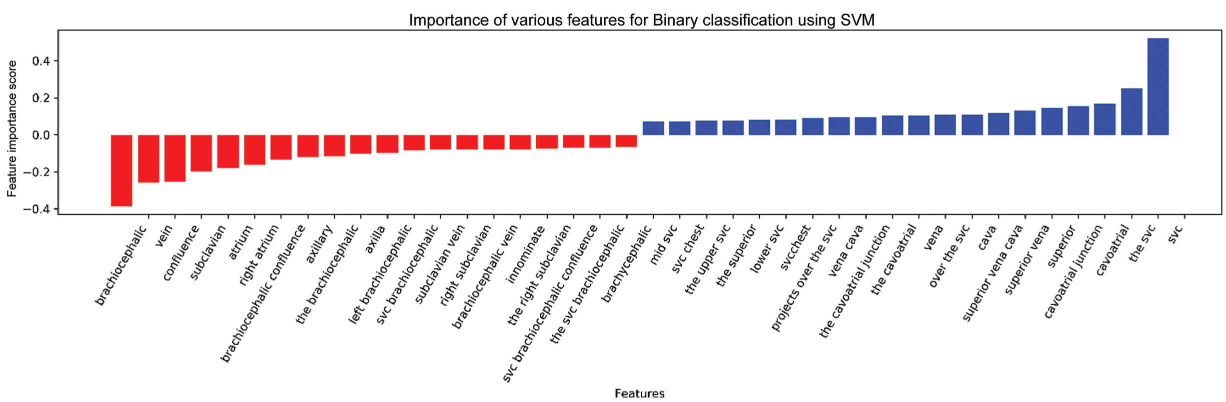

Fig. 4 Weight of importance of phrases as features for binary classification using support vector machine (SVM).

predictors of central placement, while "SVC" in combination with a word like "brachiocephalic" was a strong predictor of noncentral placement. We did not apply preprocessing on the text (e.g., removing stop words) to preserve contextual information which accounts for phrases such as "the svc," and "svc," both appearing separately.

The advantage of ML over text analysis with regular expression is demonstrated in some of the other features; for example, a phrase that had "svc" in it such as "projects over the SVC" was a good predictor of "central" classification, but "svc" in combination with other words and phrases such as "svc brachiocephalic confluence" was a good predictor of "non-central" classification. In the latter case, triggering on the term "svc" alone for central classification would result in false negative.

Our neural network model obtained 98% accuracy in binary classification and obtained 95% accuracy in predicting 12 categories. Our final model has three hidden layers with 50, 100, and 50 units, respectively (►**Fig. 5**), and the

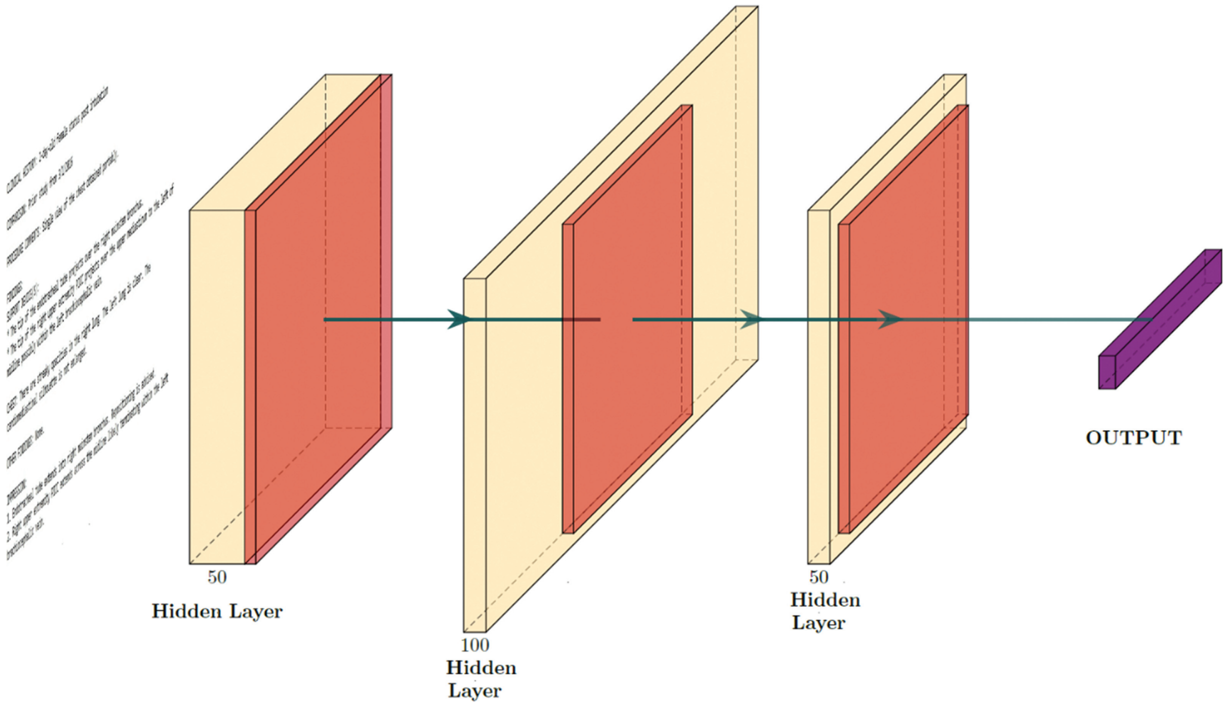

Fig. 5 Diagram of the final neural network used for classification.

hyperparameters were chosen to maximize the final classification accuracy (see ►**Supplementary Table S2** [available in the online version] for a range of hyperparameters tested for SVM and neural network classifiers). Neural network–based classifier learns the latent representations of the word embeddings using hierarchical features aggregation via the stacked hidden layers. Thus, in contrast to the SVM model, the neural network utilizes nonlinear feature combinations that are not amenable for explaining which features correspond to final accuracy.

Discussion

Timely PICC tip position detection is an important clinical issue in the NICU. Traditionally, radiograph reports are manually analyzed by the clinician to identify PICC tip location, a tedious and error-prone process. Based on recent advancements in NLP and ML, we hypothesized that such techniques could help automatically extract the anatomical location of PICC tips from unstructured radiology reports. Our experimental results, based on 17,337 reports from chest radiographs of infants with upper extremity PICCs, show that NLP ML models can correctly identify PICC tip position from radiograph reports with accuracy comparable to human classifiers.

Our study further confirms that NLP can be used successfully to extract meaningful and clinically relevant data from free-text radiology reports.[7–10,16,17,27,30] To our knowledge, this is the first time that this has been used in neonates and for PICC locations. For our approach, we were able to achieve high accuracy despite minimal preprocessing which is a promising sign for future work in this area. We believe that this work can be extended to other types of free text to extract information, such as the presence of disease processes or other important findings, to aid with patient care and research.

In our hospital, we plan to implement optimally trained ML models to automatically generate a daily report of all PICC tip locations. Our unit has a dedicated PICC team that records details on each of the active PICC lines in our unit, and they will be able to use this report to help identify when the anatomical location has migrated and trigger timely intervention when required. We have a morning huddle where leaders from the entire team (physicians, charge nurses, pharmacists, and social workers) gather, and important anticipated events or patient concerns are discussed (e.g., admissions, discharges, and expected acute events). This may be an ideal point to integrate this system into our workflow. We also eventually hope to create a reverse feed into the EHR to record the data in a structured format where it can be verified and used for clinical decision support and documentation. We have developed a web site (*https://www.picclocation.com*) that demonstrates our algorithm's performance in analyzing a new chunk of text (►**Supplementary Figure S1**, available in the online version).

A significant percentage of EHRs may contain errors or inaccuracies,[31,32] and there is evidence that EHRs result in a higher percentage of inaccurate notes than paper charts.[33] Aside from causing direct harm,[34] as well as potential legal ramifications that may stem from having inaccurate notes, the amount of time spent on documentation within EHRs has been directly associated with physician burnout,[35] and implementation of EHRs often leads to an increase in documentation time.[36] Decreasing the cognitive workload of the provider, making the interaction with EHRs more efficient, and preventing errors are some of the most basic principles of optimum EHR design. Automating as much of the work as reasonable is important in achieving these goals.[37] As future research, now that the PICC tip locations are categorized on 17,337 chest radiographs, we plan to apply image processing techniques to the radiograph images themselves to see if the PICC tip location can be predicted directly from the images. Since we can now reliably extract the PICC tip location from radiograph reports, we will be further able to expand this dataset with additional radiographs prospectively, using the trained NLP algorithms to generate labels for image-based classification. Finally, these ML-based techniques can be generalized to other uses, such as extracting pathological conditions (nodules, presence of pneumonia, pneumothorax, etc.) for both clinical and research purposes.

Limitations

Despite the study showing high accuracy, there are some limitations to our work. For instance, the chest radio graph reports at our institution all have a format with a dedicated section for support devices such as PICC lines. This standardization may limit the generalizability of our finding to sites without a similar section. Further studies could add a cohort from a different pediatric system to build on our initial algorithms and return a more generalizable result. We also excluded patients with multiple vascular cannulae which likely decreased the complexity of the reports and simplified the analyses. Further research may investigate the effectiveness of these models to handle multiple instances of catheters; however, due to the contextual nature, it would likely be a much more difficult NLP problem.

A common example of a misclassified report was when a single report described multiple images taken in succession, such as during placement of a PICC. Our classifiers frequently misclassified these reports as they contain many anatomical locations where a tip might be inappropriate, along with anatomical locations where it would be appropriate. Future work could focus on these rare report structures to help the algorithm recognize important clarifiers like, "On the final image…."

Conclusion

In this study, we have demonstrated that NLP and ML techniques can accurately extract PICC tip positions from a large set of unstructured text in infant radiograph reports. In both binary (PICC tip in appropriate as opposed to inappropriate position) and multiclass (specific anatomical location) predictions, support vector machine and neural network models obtained top accuracies. This automated process extracts categorical, structured data that can drive clinical decision support or generate labeled data for further research focused on applying ML techniques to radiographs.

Clinical Relevance Statement

This research demonstrates that support vector machines and neural networks, two machine learning (ML)-based natural language processing (NLP) algorithms, were best be able to extract peripherally inserted central catheter (PICC) tip locations from clinical radiograph reports. These algorithms can be implemented into live clinical workflows, driving decision support, and quality improvement initiatives, leading to better patient safety.

Multiple Choice Questions

1. Effective natural language processing of neonatal radiographs may most directly improve patient safety by:
   a. Producing baseline data for research projects
   b. Identifying when the reported location of an indwelling catheter changes
   c. Assisting radiologists in being more precise in their reports
   d. Determining which patients are ready for discharge

   Correct Answer: The correct answer is option b, identifying when the reported location of an indwelling catheter changes. Natural language processing (NLP) must use the information contained in the textual report. As such, it may be useful in identifying changes from one report to another. NLP is anticipated to have little effect on assisting radiologists in being more precise in their reports. An implementation of an image analysis algorithm may be able to assist radiologist with precision. While NLP may assist with producing baseline data for further research, this would not directly improve patient safety. NLP could help to determine characteristics that might indicate a patient is ready for discharge, but those textual data would likely come from progress notes or other detailed notes.

2. Each of the following artificial intelligence techniques may be useful for textual analysis except:
   a. Neural networks
   b. Support vector machines
   c. Morphological processing
   d. Random forest

   Correct Answer: The correct answer is option c, morphological processing. Morphological processing is an image processing technique that can help highlight shapes or patterns in an image, such as a circle, square, horizontal edge, or vertical edge. Identifying such shapes and their relations to each other can help extract meaning from an image. Morphological processing is not a technique used in natural language processing.[38] Neural networks comprise layers of logistic regression models to learn nonlinear patterns among features.[39] Support vector machines with polynomial and radial basis function kernels construct hyperplanes in linear and nonlinear feature spaces to classify binary outcomes.[40] Random forests use a multitude of decision trees to learn a highly irregular combination of features.[41] All the classifiers have been applied to analyze different disparate data types, including both free text and imaging data.

Protection of Human and Animal Subjects

The study was reviewed by the Institutional Review Board of Cincinnati Children's Hospital Medical Center (IRB no.: 2019–1057) and deemed exempt.

Conflict of Interest

None declared.

Acknowledgment

We would like to acknowledge and give special thanks to Tim O'Connor and Jen Regan from the Cincinnati Children's Radiology Informatics Group for their assistance in performing radiograph queries.

References

1 Sharpe E, Pettit J, Ellsbury DL. A national survey of neonatal peripherally inserted central catheter (PICC) practices. Adv Neonatal Care 2013;13(01):55–74
2 Gnannt R, Connolly BL, Parra DA, Amaral J, Moineddin R, Thakor AS. Variables decreasing tip movement of peripherally inserted central catheters in pediatric patients. Pediatr Radiol 2016;46 (11):1532–1538
3 Loskutav A, Wible BC. Peripherally inserted central catheters. In: Wible BC, ed. Diagnostic Imaging: Interventional Procedures. 2nd ed. Philadelphia, PA: Elsevier; 2018:102–107
4 Dhillon SS, Connolly B, Shearkhani O, Brown M, Hamilton R. Arrhythmias in children with peripherally inserted central catheters (PICCs). Pediatr Cardiol 2020;41(02):407–413
5 Goldwasser B, Baia C, Kim M, Taragin BH, Angert RM. Non-central peripherally inserted central catheters in neonatal intensive care: complication rates and longevity of catheters relative to tip position. Pediatr Radiol 2017;47(12):1676–1681
6 Jain A, Deshpande P, Shah P. Peripherally inserted central catheter tip position and risk of associated complications in neonates. J Perinatol 2013;33(04):307–312
7 Trivedi G, Dadashzadeh ER, Handzel RM, Chapman WW, Visweswaran S, Hochheiser H. Interactive NLP in clinical care: identifying incidental findings in radiology reports. Appl Clin Inform 2019;10 (04):655–669
8 Jones BE, South BR, Shao Y, et al. Development and validation of a natural language processing tool to identify patients treated for pneumonia across VA emergency departments. Appl Clin Inform 2018;9(01):122–128
9 Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016;279(02):329–343
10 Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform 2017;73:14–29
11 Goldberg Y, Hirst G. Neural network methods in natural language processing. Synthesis Lectures on Human Language Technologies 2017;10(01):1–309
12 Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. JMIR Med Inform 2020;8(01):e16023
13 Nawab K, Ramsey G, Schreiber R. Natural language processing to extract meaningful information from patient experience feedback. Appl Clin Inform 2020;11(02):242–252
14 Zhu VJ, Walker TD, Warren RW, Jenny PB, Meystre S, Lenert LA. Identifying falls risk screenings not documented with administrative codes using natural language processing. AMIA Annu Symp Proc 2018;2017:1923–1930
15 Nguyen AN, Truran D, Kemp M, et al. Computer-assisted diagnostic coding: effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings. AMIA Annu Symp Proc 2018; 2018:807–816
16 Grundmeier RW, Masino AJ, Casper TC, et al; Pediatric Emergency Care Applied Research Network. Identification of long bone fractures in radiology reports using natural language processing to support healthcare quality improvement. Appl Clin Inform 2016;7(04):1051–1068
17 Sevenster M, Buurman J, Liu P, Peters JF, Chang PJ. Natural language processing techniques for extracting and categorizing finding measurements in narrative radiology reports. Appl Clin Inform 2015;6(03):600–110
18 Malmasi S, Sandor NL, Hosomura N, Goldberg M, Skentzos S, Turchin A. Canary: an NLP platform for clinicians and researchers. Appl Clin Inform 2017;8(02):447–453
19 Deo RC. Machine learning in medicine. Circulation 2015;132(20): 1920–1930
20 Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform 2018;85:30–39
21 Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. Radiology 2018;288(02):318–328
22 McBee MP, Awan OA, Colucci AT, et al. Deep learning in radiology. Acad Radiol 2018;25(11):1472–1480
23 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542(7639):115–118
24 Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA 2019;322(18):1806–1816
25 Dominiczak J, Khansa L. Principles of automation for patient safety in intensive care: learning from aviation. Jt Comm J Qual Patient Saf 2018;44(06):366–371
26 Lee H, Mansouri M, Tajmir S, Lev MH, Do S. A deep-learning system for fully-automated peripherally inserted central catheter (PICC) tip detection. J Digit Imaging 2018;31(04):393–402
27 Kim C, Zhu V, Obeid J, Lenert L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. PLoS One 2019;14(02):e0212778
28 Rossum GV, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009
29 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–2830
30 Blagev DP, Lloyd JF, Conner K, et al. Follow-up of incidental pulmonary nodules and the radiology report. J Am Coll Radiol 2016;13(2, suppl):R18–R24
31 Bell SK, Delbanco T, Elmore JG, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020;3(06):e205867
32 Lam BD, Bourgeois F, Dong ZJ, Bell SK. Speaking up about patient-perceived serious visit note errors: patient and family experiences and recommendations. J Am Med Inform Assoc 2021;28 (04):685–694
33 Yadav S, Kazanji NKC N. et al. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. J Am Med Inform Assoc 2017;24(01):140–144
34 Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. JAMA 2018;319(12):1276–1278
35 Robertson SL, Robinson MD, Reid A. Electronic health record effects on work-life balance and burnout within the I^3 population collaborative. J Grad Med Educ 2017;9(04):479–484
36 Carayon P, Wetterneck TB, Alyousef B, et al. Impact of electronic health record technology on the work and workflow of physicians in the intensive care unit. Int J Med Inform 2015;84(08):578–594
37 Zahabi M, Kaber DB, Swangnetr M. Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation. Hum Factors 2015;57(05): 805–834
38 Bishop CM. Pattern Recognition and Machine Learning. New York, NY: Springer; 2016
39 Aggarwal CC. Neural Networks and Deep Learning: A Textbook. Switzerland: Springer; 2018
40 Shawe-Taylor JCN. Kernel Methods for Pattern Analysis. Davis, CA: Cambridge University Press; 2004
41 Breiman L. Random forests. Mach Learn 2001;45(01):5–32

