

RESEARCH

Open Access



ClotCatcher: a novel natural language model to accurately adjudicate venous thromboembolism from radiology reports

Jeffrey Wang^{1*}, Joao Souza de Vale¹, Saransh Gupta¹, Pulakesh Upadhyaya¹, Felipe A. Lisboa^{2,3,4}, Seth A. Schobel^{2,3,4}, Eric A. Elster^{2,3}, Christopher J. Dente^{2,5,6}, Timothy G. Buchman^{2,5,7} and Rishikesan Kamaleswaran^{1,2}

Abstract

Introduction Accurate identification of venous thromboembolism (VTE) is critical to develop replicable epidemiological studies and rigorous predictions models. Traditionally, VTE studies have relied on international classification of diseases (ICD) codes which are inaccurate – leading to misclassification bias. Here, we developed ClotCatcher, a novel deep learning model that uses natural language processing to detect VTE from radiology reports.

Methods Radiology reports to detect VTE were obtained from patients admitted to Emory University Hospital (EUH) and Grady Memorial Hospital (GMH). Data augmentation was performed using the Google PEGASUS paraphraser. This data was then used to fine-tune ClotCatcher, a novel deep learning model. ClotCatcher was validated on both the EUH dataset alone and GMH dataset alone.

Results The dataset contained 1358 studies from EUH and 915 studies from GMH ($n = 2273$). The dataset contained 1506 ultrasound studies with 528 (35.1%) studies positive for VTE, and 767 CT studies with 91 (11.9%) positive for VTE. When validated on the EUH dataset, ClotCatcher performed best (AUC = 0.980) when trained on both EUH and GMH dataset without paraphrasing. When validated on the GMH dataset, ClotCatcher performed best (AUC = 0.995) when trained on both EUH and GMH dataset with paraphrasing.

Conclusion ClotCatcher, a novel deep learning model with data augmentation rapidly and accurately adjudicated the presence of VTE from radiology reports. Applying ClotCatcher to large databases would allow for rapid and accurate adjudication of incident VTE. This would reduce misclassification bias and form the foundation for future studies to estimate individual risk for patient to develop incident VTE.

Keywords Natural language processing, Machine learning, Venous thromboembolism

*Correspondence:

Jeffrey Wang

jwang37@emory.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Venous thromboembolism (VTE) is defined as the development of either a deep venous thrombus (DVT) or pulmonary embolism (PE) and is widely considered to be a preventable and leading cause of death worldwide [1–3]. VTE is estimated to occur in 1 in 1000 patients in the United States [4] and is associated with increased cost [5], hospital length of stay [6], morbidity [7], and a higher risk of both short-term and long-term mortality [8]. There have been several nation-wide campaigns addressing VTE as a public health issue, including a call in 2008 to prevent in-hospital VTE by the United States Surgeon General [9, 10].

Several studies utilizing large electronic medical record databases to determine predictors and outcomes in patients who develop VTE have recently been published; however, they are limited by their reliance on using International Classification of Disease (ICD) codes to determine the incidence of VTE [11–13]. While ICD codes are easily accessible and can be rapidly applied to large databases, they have several important limitations. First, their accuracy has been called into question with recent analyses finding that among patients with an ICD diagnosis of VTE, only 30%–60% have clinical documentation from radiological reports to support this diagnosis [12, 14, 15]. Another important limitation is that ICD codes do not provide information regarding the time of incident VTE, as it can only be used to identify whether the condition was present or absent. Without data on timing, analysis and modeling are limited to more traditional and simplistic approaches such as logistic regression.

To address this, recent approaches have used Natural Language Processing (NLP), a field within machine learning that concentrates on text data with the goal of developing models that can understand human writing. Using NLP, researchers have used models that can identify whether VTE is present or absent from clinical notes (see Supplemental Table 1) [16–19]. These approaches have utilized tools that are rules based or advanced text miners; however, a major limitation in most of these approaches is that they use tools that are considered black box in which the architecture of the tools is not available. Some of these tools were from a commercialized source, further limiting the ability to replicate the methods. While the architecture in the tool published by Verma et al. is available, the tool used in this study used a rules-based NLP tool, rather than newer available tools [19].

In this investigation we develop ClotCatcher, which is a novel deep learning technology that incorporates the use of BERT (Bidirectional Encoder Representations

from Transformers). BERT is a state-of-the-art model that incorporates several key advancements in the field of NLP that can be applied to textual data. First, BERT uses a transformer, which is a neural network used predominantly in NLP that allows the tool to weight certain parts of the sentence to predict the relationships between the words in the sentence. Having an improved understanding of the sentence allows the encoder to better detect key words allowing the tool to interpret hidden representation of the text, resulting in increased accuracy when decoding. These advancements led to the development of BERT, which achieved state-of-the-art results in NLP tasks. As a result, BERT has been utilized as a pre-trained base for constructing models that cater to specific contexts, such as biomedical texts, scientific publications, clinical notes, and patient information [20–22].

Secondly, BERT is also the first to use bidirectional models, in comparison to earlier models that trained unidirectionally, meaning that the tool only trained on language going in one direction (i.e., left-to-right). Bidirectional models, such as BERT, utilize a novel methodology by allowing input data to be input from either direction (i.e., both left-to-right and right-to-left), improving contextual understanding. Finally, BERT uses Masked Language Modeling (MLM), which is an approach to further improve prediction and contextual understanding by randomly masking words in the text and forcing the model to predict the word that is randomly masked. This allows the NLP tool to improve the contextual understanding of the text by looking at the surrounding sentence [23].

The objectives of our study were (1) to develop a deep learning tool using a deep learning model to detect VTE from clinical reports, (2) create a tool using open source methodologies that would not rely on black box algorithms, allowing for reproducibility, and (3) create a generalizable NLP tool by training the tool on datasets from two different hospital systems.

Methods

Study design

This study used a retrospective, observational design to create and validate ClotCatcher, a novel deep learning model to accurately adjudicate the presence or absence of venous thromboembolism (VTE) from radiology notes. This study was approved by the Emory University Institutional Review Board (STUDY00000302) under waiver of consent due to the retrospective nature of the study. This investigation was carried out in accordance to the Emory University Institutional Review Board guidelines and regulations.

Data source

We included all radiology studies from patients who were admitted to either Emory University Hospital (EUH) (years 2014 – 2021) or Grady Memorial Hospital (GMH) (years 2014 – 2022). During the years noted, EUH used PowerChart™ from Cerner™ and GMH used Epic™. We selected only radiology studies that are used to evaluate for VTE (see Table 1). We did not include ventilation-perfusion scintigraphy (i.e. V:Q scans) as the results are given as a probability. We then randomly selected 5% of all studies from GMH and 2.5% of studies from EUH. The reports of all radiological studies were extracted into a document and a physician (JW) adjudicated whether VTE was present or absent. Notes which were not finalized or were incomplete were excluded from the study (n = 5).

Table 1 List of imaging studies selected

Imaging Studies Selected
Ultrasound – Bilateral Upper Extremity
Ultrasound – Right Lower Extremity
Ultrasound – Left Lower Extremity
Ultrasound – Bilateral Lower Extremity
Ultrasound – Right Lower Extremity
Ultrasound – Left Lower Extremity
Computed Tomography Chest w/ Intravenous Contrast (Pulmonary angiography)

Cleaning and creating training dataset

The radiology reports were first extracted, and all text converted to lowercase to standardize the text. The reports from both GMH and EUH were then randomly split 80% to 20% (training to validation) for both hospitals. Randomly splitting the dataset into training and validation helps the model avoid overfitting, where the model performs well on the training set but poorly on new, unseen data. To maximize the potential from the training dataset, we performed data augmentation using paraphrasing, a technique that generates new versions of the text using different words and/or syntax, while still preserving the meaning of the original text. Two distinct strategies were developed, one with paraphrasing and one without paraphrasing. Paraphrasing can be done manually, however in our dataset, the Google PEGASUS model, a transformer-based neural network-based NLP tool was used which automates data paraphrasing. After applying the Google PEGASUS model to the training dataset, we generated 20 additional studies for each unique study, resulting in 21 × amplification of the training dataset.

While we started with datasets from EUH and GMH, we created a third dataset by combining EUH and GMH resulting in three unique combinations – EUH alone, GMH alone, and EUH combined with GMH (see Fig. 1 for training dataset creation and model validation pipeline). We created six total training datasets, the initial three used the datasets without applying paraphrasing.

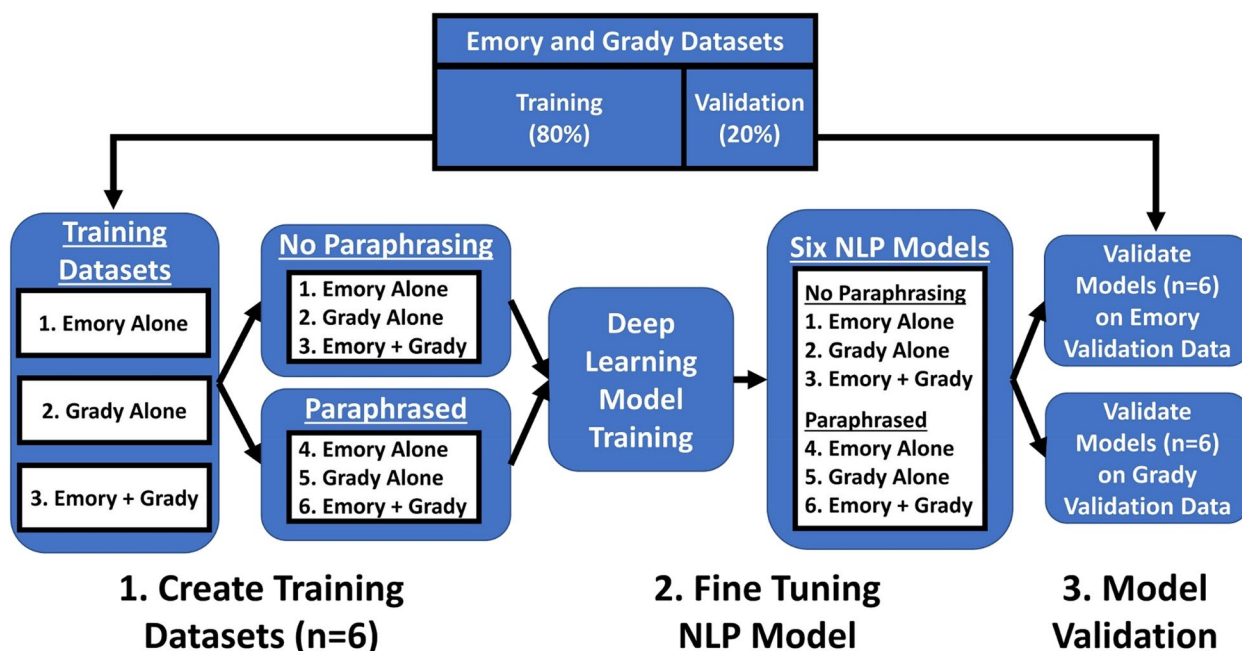


Fig. 1 Overview of the architecture of the natural language processing pipeline

The additional three were created after using data augmentation by applying paraphrasing, resulting in datasets that were $21 \times$ the size of the original datasets [24].

Clotcatcher tool development

The training datasets were then used to create the ClotCatcher tool by fine-tuning the BERT model to better adjudicate for the presence or absence of VTE based on the text available within the radiology report. The physician adjudicated result was considered the gold standard. Comparing the paraphrased dataset to the datasets not utilizing paraphrasing allowed us to evaluate the impact of data augmentation.

To evaluate the generalizability of ClotCatcher, we trained six deep learning models from six datasets (see Fig. 1). The first three were 1) EUH alone (without paraphrasing), 2) GMH alone (without paraphrasing), and 3) EUH combined with GMH (without paraphrasing). The final three datasets were the same as above, however with paraphrasing applied for data augmentation.

These six training datasets were then used to fine-tune the ClotCatcher tool to create six different deep learning models which were then validated on the EUH dataset alone and the GMH dataset alone. Key metrics to measure model performance such as sensitivity, specificity, F1 score, accuracy, and area under the receiver operating curve (AUC) were determined for each model and validation cohort combination. The architecture of the ClotCatcher tool can be seen in Fig. 1. All program codes were written for Python programming language (Version 3.9).

All ICD-9 and ICD-10 codes were extracted from each hospitalization for which the imaging report originated. ICD codes consistent with VTE were further determined based on previous literature [25]. Hospitalizations with ICD codes consistent with VTE were adjudicated as VTE positive by ICD Code. This was compared against physician adjudication, which is considered the gold standard. Key metrics to measure model performance such as sensitivity, specificity, F1 score, accuracy, and area under the receiver operating curve (AUC) were determined for GMH and EUH.

Results

Analytic cohort

After applying appropriate inclusion criteria for radiological studies, we obtained 1358 studies from EUH and 915 studies from GMH (see Fig. 2). Both EUH and GMH datasets were randomly split into derivation (EUH $n=1086$, GMH $n=732$) and validation (EUH $n=272$, GMH $n=183$). The baseline characteristics for the patients stratified by hospital are listed in Table 2. Patients from GMH were younger (53.5 ± 16.9 vs

57.8 ± 17.5 years), less likely to be female (381/895 [32.5%] vs 754/1342 [56.2%]), and more likely to self-identify as African American (718 [80%] vs 613 [45.7%]). The proportion of ultrasound studies adjudicated to have VTE by a physician was 177/653 (27.1%) at EUH and 351/853 (41.1%) at GMH. The proportion of CT studies adjudicated to have VTE by a physician was 88/705 (12.5%) at EUH and 3/62 (4.8%) at GMH (Table 3). The median time to study ordered was 33.7 [Interquartile Range (IQR): 14.4, 139.2] hours, and 9.8 [IQR: 4.0, 93.6] hours for GMH (Table 2). When stratified by whether the study was positive for VTE (1) or not (2), the distribution plots are presented in Supplemental Fig. 3.

ClotCatcher tool validation

We performed validation of all six deep learning models generated from ClotCatcher on the validation datasets from EUH and GMH separately. When validated on the EUH dataset, all six models produced excellent AUCs, ranging from 0.966 to 0.980 (Fig. 3, Table 4). Interestingly, the model with the highest AUC validated on the EUH dataset was fine-tuned on EUH combined with GMH data without paraphrasing (AUC 0.980, 95% Confidence Interval [CI]: 0.977 – 0.983). When validated on the GMH dataset, all six models also had excellent AUCs ranging from 0.988 – 0.995 (Fig. 4, Table 5). The model with the highest AUC validated on the GMH dataset was fine-tuned on EUH combined with GMH data utilizing paraphrasing for data augmentation. ClotCatcher performed better than using ICD to adjudicate whether VTE was present or not during the hospitalization across all metrics (Supplemental Table 2).

Discussion

In this study, we created ClotCatcher, a deep learning tool that was able to accurately and rapidly adjudicate the presence or absence of VTE from text in radiological reports. We obtained excellent results with our best models demonstrating an AUC of 0.980 on the EUH dataset and an AUC of 0.995 on the GMH dataset. Our approach is novel as it is the first study to use either BERT – which is considered state-of-the-art technology in natural language processing or the paraphrasing technique in developing a tool to detect VTE from radiology studies. Furthermore, the architecture of the ClotCatcher tool uses non-proprietary tools that are readily available. Previous approaches have relied on black box or commercialized tools that are not easily replicated. By combining these techniques into our pipeline, we have created a tool which is both powerful and replicable.

The high metrics of the ClotCatcher tool is in large part due to its use of BERT, which has been established as the industry standard in NLP and has found

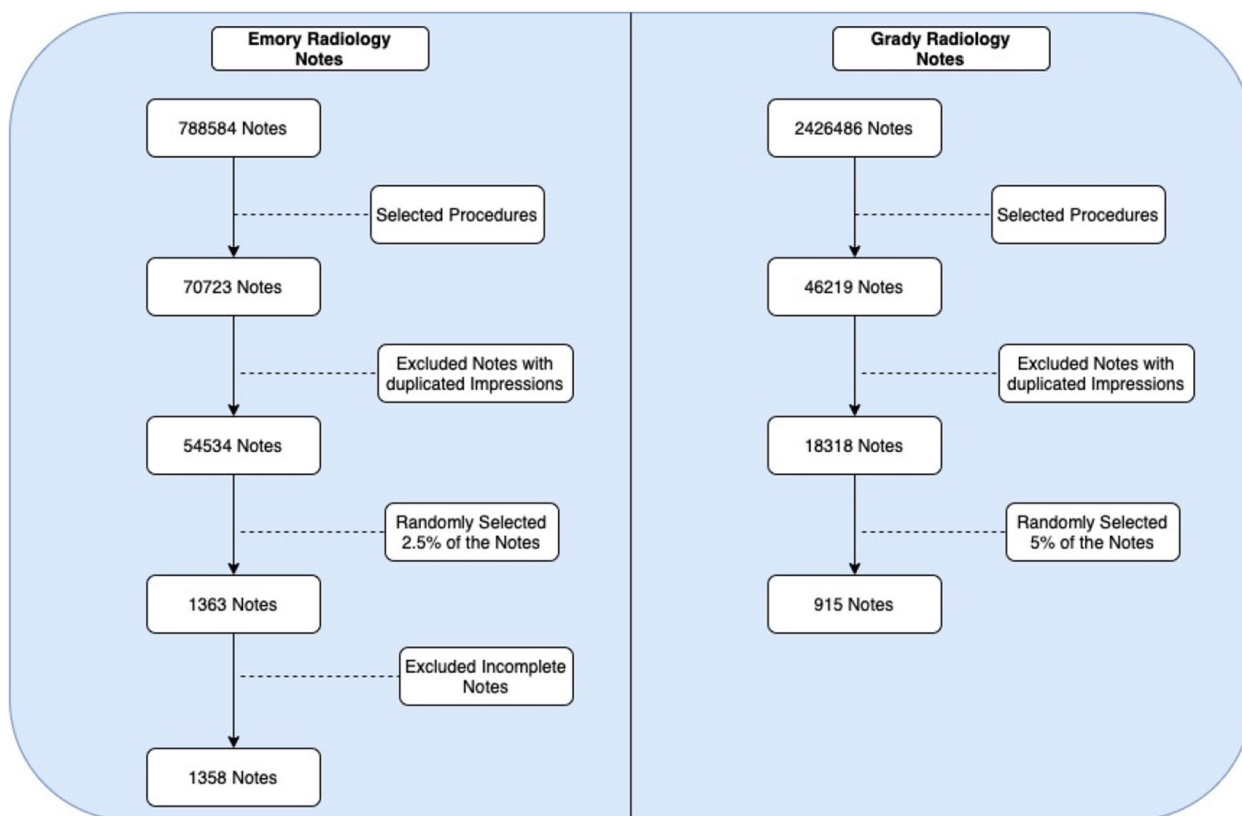


Fig. 2 CONSORT diagram describing analytic cohort for studies

Table 2 Baseline characteristics for analytic cohort by hospital

	Emory University Hospital (n = 1342)	Grady Memorial Hospital (n = 895)
Age	57.8 ± 17.5	53.5 ± 16.9
Female, n (%)	754 (56.2%)	381 (32.5%)
Race		
Caucasian	643 (47.9%)	83 (9.3%)
African American	613 (45.7%)	718 (80%)
Asian	35 (2.6%)	8 (0.9%)
History of Diabetes	427 (31.8%)	112 (12.5%)
History of MI	183 (13.6%)	29 (3.2%)
History of CHF	407 (30.3%)	100 (11.1%)
History of PVD	205 (15.3%)	21 (2.3%)
Chronic Pulmonary Disease	448 (33.4%)	117 (13%)
CKD or Dialysis	340 (25.3%)	96 (10%)
History of Cancer	246 (18.3%)	20 (2.2%)
HIV/AIDS	25 (1.9%)	17 (1.9%)
Time from Admission to Radiological Study (hours)	33.7 [14.4 – 39.2]	9.8 [4.0 – 93.6]

widespread use from medical research to Google searches. In our analysis, we used BioBERT, the form of BERT pre-trained on biological data, which is the

first use of BERT in adjudicating VTE from radiological studies. The significance of these tools are readily apparent when considering that only 2.5% of the Emory dataset and 5% of the Grady dataset was clinician adjudicated. Using ClotCatcher, we can now accurately and rapidly adjudicate all available studies within the EUH and GMH database.

There were important differences in the GMH and EUH patient populations (Table 2). These differences can be explained by the hospital locations and the communities which they serve. Patients at GMH were younger, more likely to be male, and had a higher proportion self-identify as African American. As patients are younger at GMH, younger patients generally have less comorbidities resulting in a lower prevalence of ICD codes in this population. Secondly, there are differences in VTE evaluation, at EUH, 48% (653/1358) of studies ordered were ultrasound, whereas at GMH, this was 93% (853/915). Secondly, while there are no readily apparent clinical differences to explain differences in ordering pattern, it is possible that differences in physician ordering practice could be driving these differences. In a retrospective observational study of emergency room physicians ordering CT studies to evaluate for VTE, the number of studies ordered by

Table 3 Radiological study count and proportion positive for venous thromboembolism

	Emory (n = 1358)	Grady (n = 915)	Total (n = 2273)
Ultrasound	653	853	1506
- Proportion Positive for VTE	177 (27.1%)	351 (41.1%)	528 (35.1%)
CT Studies	705	62	767
- Proportion Positive for VTE	88 (12.5%)	3 (4.8%)	91 (11.9%)

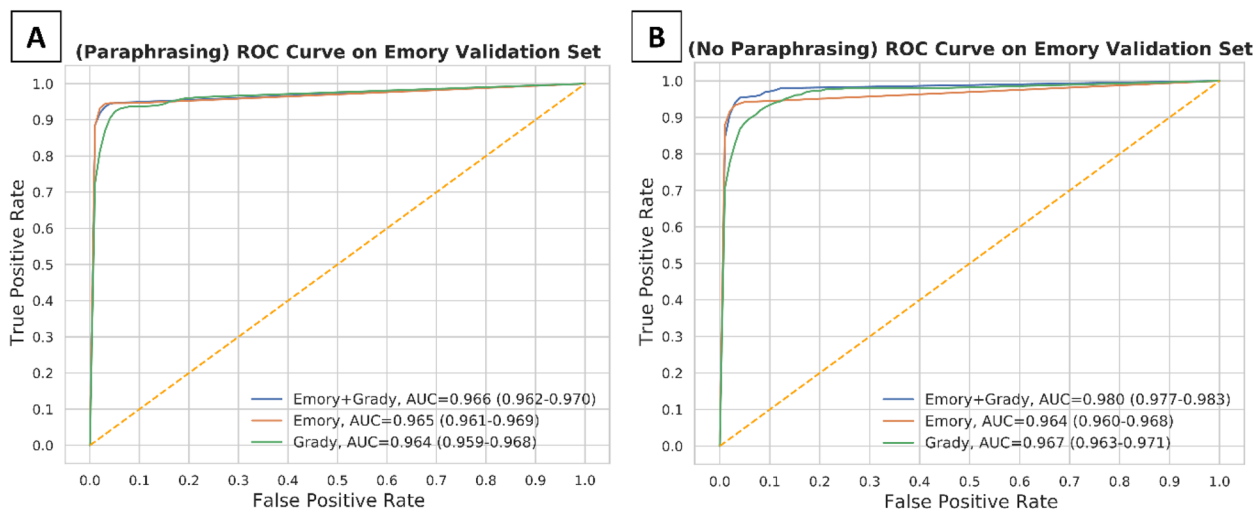


Fig. 3 Model validation on the Emory dataset after training on either 1) Emory and Grady, 2) Emory alone, or 3) Grady alone dataset. This was evaluated with (3A) paraphrasing and (3B) without

Table 4 The sensitivity, specificity, F1 Score, accuracy and the area under the receiver operation curve (AUC) for the model validated on the Emory dataset with and without paraphrasing

Training Data	Emory Validation Set				
	Sensitivity	Specificity	F1 Score	Accuracy	AUC score
Emory + Grady (Paraphrased)	0.887	0.990	0.922	0.971	0.966
Emory (Paraphrased)	0.943	0.986	0.943	0.978	0.965
Grady (Paraphrased)	0.717	0.991	0.817	0.938	0.964
Emory + Grady	0.830	0.986	0.880	0.956	0.980
Emory	0.849	0.990	0.900	0.963	0.964
Grady	0.811	0.978	0.851	0.945	0.967

Best in class metrics are bolded in red

each physician varied greatly from 25 to 141 CT studies per physician [26].

Thirdly, the CT studies at GMH have a low positivity rate (4.8%) compared to EUH (12.5%). This can be in part explained that GMH is the only level one trauma center in metropolitan Atlanta. During the initial evaluation of complex poly-trauma, CT studies are often obtained to evaluate multiple pathologies simultaneously. Therefore,

the pre-test probability for VTE in these situations are low, which can partially explain the low positivity rate. This seemingly low positive rate is consistent with existing literature demonstrating that among emergency room physicians ordering CT studies to evaluate for VTE, the proportion positive was 6.9% [26].

The tool directly addresses the inaccuracies from misclassification bias due to using ICD codes in studies

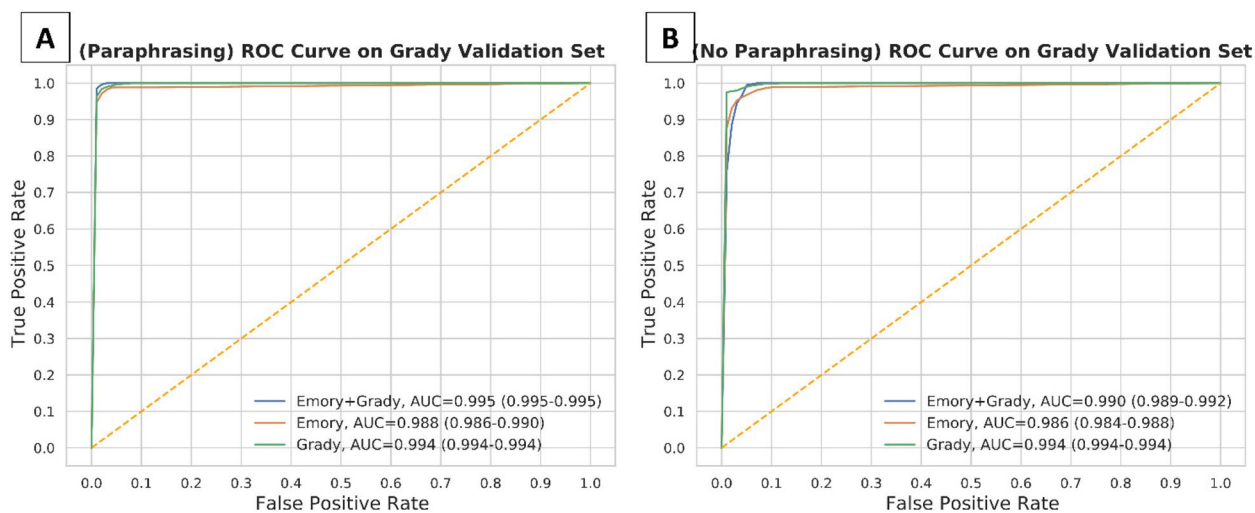


Fig. 4 Model validation on the Grady dataset after training on either 1) Emory and Grady, 2) Emory alone, or 3) Grady alone dataset. This was evaluated with (4A) paraphrasing and (4B) without

Table 5 The sensitivity, specificity, F1 Score, accuracy and the area under the receiver operation curve (AUC) for the model validated on the Grady dataset with and without paraphrasing

Training Data	Grady Validation Set				
	Sensitivity	Specificity	F1 Score	Accuracy	AUC score
Emory + Grady (Paraphrased)	0.984	0.992	0.984	0.990	0.995
Emory (Paraphrased)	0.968	0.984	0.968	0.978	0.988
Grady (Paraphrased)	0.984	0.992	0.984	0.990	0.994
Emory + Grady	0.968	0.975	0.960	0.973	0.990
Emory	0.903	0.984	0.933	0.957	0.986
Grady	0.968	0.975	0.960	0.973	0.994

Best in class metrics are bolded in red

investigating VTE. ICD codes are known to be inaccurate and previous literature demonstrated that the positive predictive value of a VTE ICD code to predict the presence of VTE diagnosed during hospitalization was poor, around 50%. By applying tools such as ClotCatcher to radiological studies, investigators will be able to accurately adjudicate the presence of VTE use radiological studies. This will allow for more accurate epidemiological studies and reduce misclassification bias.

In addition to accurate adjudication, all radiological studies have an associated time stamp which will provide the time VTE was diagnosed on a radiological study. When relying on traditional methods which use ICD codes, lack of time data limited analyses to more simplistic methods such as logistic regression. Incorporating the time of the event will allow for more sophisticated analyses such as Cox-proportional hazard modeling and

Kaplan Meier analysis. Furthermore, these modeling approaches will form the basis to create automated individualized risk prediction scores for patients admitted to the hospital. This is particularly relevant for patients who have their chemoprophylaxis for VTE held as applying individualized risk prediction models can alert clinicians to consider starting chemoprophylaxis in patients with a high risk for in-hospital VTE.

An important aspect of applying machine learning models to healthcare data is the concept of generalization, which refers to the ability of a model to accurately predict outcomes from data that is different from the source population it was originally trained on. In healthcare, data limitations often constrain researchers to use data from the same hospital for training and testing; however, given the increasing popularity of machine learning models in healthcare, it is crucial to investigate the

performance of models on data sourced from different hospitals and healthcare systems [27, 28]. Our model had excellent performance on both the EUH and the GMH datasets, demonstrating that our model is not limited by the origin of the training data. Furthermore, the two hospitals used different electronic medical records during the study period which further supports the generalizability of our model. The slight variation in the metrics could be due to different documentation standards across hospitals. Future directions would include replicating our analysis using additional hospital systems.

There were several limitations. We used a subset of the available data, given limitations in time and resources in utilizing physician adjudication of the radiological studies. We were also limited to studies that were commonly used to identify VTE, thereby missing incidental VTE found by other studies (i.e., portal vein thrombus, splanchnic vein thrombus, etc.). We also excluded ventilation-perfusion scans from this study given that the interpretation of these studies is usually provided as probabilities. Finally, radiologists at GMH can have appointments at EUH as well and therefore there may be similarities between these two hospitals in producing radiological reports.

In conclusion, we present the results from ClotCatcher, a novel deep learning tool to accurately and rapidly adjudicate the presence or absence of VTE from radiological reports. The tool can be readily replicated using existing open-source tools such as BERT and the paraphrasing technique. Validation of this ClotCatcher serves as the foundation for improving identification of VTE cases from large databases.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-023-02369-z>.

Additional file 1: Supplemental Table 1. Summary of published natural language processing models to adjudicate the presence or absence of deep vein thrombosis from radiological studies. **Supplemental Figure 1.** Model calibration on the Emory dataset after training on either 1) Emory and Grady, 2) Emory alone, or 3) Grady alone dataset. This was evaluated with (1A) paraphrasing and (1B) without. **Supplemental Figure 2.** Model calibration on the Grady dataset after training on either 1) Emory and Grady, 2) Emory alone, or 3) Grady alone dataset. This was evaluated with (1A) paraphrasing and (1B) without. **Supplemental Figure 3.** Distribution plots for Time from Admission to Radiological Study Order. **Supplemental Table 2.** Metrics for VTE positive using ICD Codes.

Acknowledgements

The authors acknowledge Kim Tierney and Elizabeth Dee for their assistance with the manuscript. The work by the authors was done as a part with funding from the National Institutes of Health; however, the opinions expressed are not necessarily those of the National Institutes of Health.

Authors' contributions

JW JD SG PU RK designed research, performed research, analyzed the taper, wrote paper. FAL SAS EAL CJD TGB RK wrote paper.

Authors' information (optional) disclosures

The contents of this presentation are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of Uniformed Services University of the Health Sciences (USUHS), The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., the Department of Defense (DoD) or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Funding

JW is supported by the GA CTSa (UL1TR002378, TL1TR002382) and the National Heart, Lung and Blood Institute of the National Institutes of Health under award number 5T32HL007745. The funding bodies played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript. R Kamaleswaran was supported by the National Institutes of Health under Award Numbers R01GM139967, R21GM151703, R21GM148931, OT2OD032701, and UL1TR002378. Research activities leading to the development of this article were funded by the Department of Defense's Defense Health Program-Joint Program Committee 6/Combat Casualty Care (USUHS HT9404-13-1-0032 and HU0001-15-2-0001).

Availability of data and materials

Data is available upon reasonable requests. For original data, please contact rkamaleswaran@emory.edu.

Declarations

Ethics approval and consent to participate

This study was approved by the Emory University Institutional Review Board (STUDY00000302). This study was approved by the Emory University Institutional Review Board under waiver of consent due to the retrospective nature of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biomedical Informatics, Emory University School of Medicine, 1462 Clifton Road, Suite 504, Atlanta, GA 30322, USA. ²Surgical Critical Care Initiative (SC2i), Uniformed Services University of the Health Sciences, Bethesda, MD 20814, USA. ³Department of Surgery, Uniformed Services University of the Health Sciences and Walter Reed National Military Medical Center, Bethesda, MD 20814, USA. ⁴Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD 20817, USA. ⁵Emory Department of Surgery, Emory University School of Medicine, Atlanta, GA, USA. ⁶Grady Memorial Hospital, Atlanta, GA, USA. ⁷Emory Critical Care Center, Atlanta, GA, USA.

Received: 14 June 2023 Accepted: 7 November 2023

Published online: 16 November 2023

References

- Clagett GP, Anderson FA Jr, Heit J, Levine MN, Wheeler HB. Prevention of Venous Thromboembolism. *Chest*. 1995;108(4):312S-334S. https://doi.org/10.1378/chest.108.4_Supplement.312S.
- Heit JA. Prevention of venous thromboembolism. *Clin Geriatr Med*. 2001;17(1):71-92. [https://doi.org/10.1016/S0749-0690\(05\)70107-5](https://doi.org/10.1016/S0749-0690(05)70107-5).
- Lau BD, Haut ER. Practices to prevent venous thromboembolism: a brief review. *BMJ Qual Saf*. 2014;23(3):187-95. <https://doi.org/10.1136/bmjqs-2012-001782>.
- Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. *Am J Prev Med*. 2010;38(4 Suppl):S495-501. <https://doi.org/10.1016/j.amepre.2009.12.017>.
- Cohoon KP, Leibson CL, Ransom JE, et al. Direct medical costs attributable to venous thromboembolism among persons hospitalized for

- major operation: a population-based longitudinal study. *Surgery*. 2015;157(3):423–31. <https://doi.org/10.1016/j.surg.2014.10.005>.
6. Correction to: Call to Action to Prevent Venous Thromboembolism in Hospitalized Patients: A Policy Statement From the American Heart Association. *Circulation*. 2021 143(7):e249–e249. <https://doi.org/10.1161/CIR.0000000000000956>.
 7. Ja HEIT. Venous thromboembolism: disease burden, outcomes and risk factors. *J Thromb Haemost*. 2005;3(8):1611–7. <https://doi.org/10.1111/j.1538-7836.2005.01415.x>.
 8. Søgaard KK, Schmidt M, Pedersen L, Horváth-Puhó E, Sørensen HT. 30-year mortality after venous thromboembolism: a population-based cohort study. *Circulation*. 2014;130(10):829–36. <https://doi.org/10.1161/circulationaha.114.009107>.
 9. Streiff MB, Brady JP, Grant AM, Grosse SD, Wong B, Popovic T. CDC Grand Rounds: preventing hospital-associated venous thromboembolism. *MMWR Morb Mortal Wkly Rep*. 2014;63(9):190–3.
 10. (US) OotSG. The Surgeon General's Call to Action to Prevent Deep Vein Thrombosis and Pulmonary Embolism. Office of the Surgeon General (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK44178/>.
 11. Neeman E, Liu V, Mishra P, et al. Trends and risk factors for venous thromboembolism among hospitalized medical patients. *JAMA Netw Open*. 2022;5(11):e2240373–e2240373. <https://doi.org/10.1001/jamanetworkopen.2022.40373>.
 12. Nelson RE, Grosse SD, Waitzman NJ, et al. Using multiple sources of data for surveillance of postoperative venous thromboembolism among surgical patients treated in Department of Veterans Affairs hospitals, 2005–2010. *Thromb Res*. 2015;135(4):636–42. <https://doi.org/10.1016/j.thromres.2015.01.026>.
 13. Boulet SL, Grosse SD, Hooper WC, Beckman MG, Atrash HK. Prevalence of venous thromboembolism among privately insured US adults. *Arch Intern Med*. 2010;170(19):1774–5. <https://doi.org/10.1001/archinternmed.2010.336>.
 14. Baumgartner C, Go AS, Fan D, et al. Administrative codes inaccurately identify recurrent venous thromboembolism: The CVRN VTE study. *Thromb Res*. 2020;189:112–8. <https://doi.org/10.1016/j.thromres.2020.02.023>.
 15. Pellathy T, Saul M, Clermont G, Dubrawski AW, Pinsky MR, Hravnak M. Accuracy of identifying hospital acquired venous thromboembolism by administrative coding: implications for big data and machine learning research. *J Clin Monit Comput*. 2022;36(2):397–405. <https://doi.org/10.1007/s10877-021-00664-6>.
 16. Woller B, Daw A, Aston V, et al. Natural language processing performance for the identification of venous thromboembolism in an integrated health-care system. *Clin Appl Thromb Hemost*. 2021;27:10760296211013108. <https://doi.org/10.1177/10760296211013108>.
 17. Gálvez JA, Pappas JM, Ahumada L, et al. The use of natural language processing on pediatric diagnostic radiology reports in the electronic health record to identify deep venous thrombosis in children. *J Thromb Thrombolysis*. 2017;44(3):281–90. <https://doi.org/10.1007/s11239-017-1532-y>.
 18. Shi J, Hurdle JF, Johnson SA, et al. Natural language processing for the surveillance of postoperative venous thromboembolism. *Surgery*. 2021;170(4):1175–82. <https://doi.org/10.1016/j.surg.2021.04.027>.
 19. Verma AA, Masoom H, Pou-Prom C, et al. Developing and validating natural language processing algorithms for radiology reports compared to ICD-10 codes for identifying venous thromboembolism in hospitalized medical patients. *Thromb Res*. 2022;209:51–8. <https://doi.org/10.1016/j.thromres.2021.11.020>.
 20. Huang K, Altosaar J, Ranganath R. Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:190405342. 2019; <https://arxiv.org/abs/1904.05342>.
 21. Lee J-S, Hsiang J. Patentbert: Patent classification with fine-tuning a pre-trained bert model. arXiv preprint arXiv:190602124. 2019; <https://arxiv.org/abs/1906.02124>.
 22. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234–40.
 23. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018; <https://arxiv.org/abs/1810.04805>.
 24. Feng SY, Gangal V, Wei J, et al. A survey of data augmentation approaches for NLP. arXiv preprint arXiv:210503075. 2021; <https://arxiv.org/abs/2105.03075>.
 25. Weller SC, Porterfield L, Davis J, Wilkinson GS, Chen L, Baillargeon J. Incidence of venous thrombotic events and events of special interest in a retrospective cohort of commercially insured US patients. *BMJ Open*. 2022;12(2):e054669. <https://doi.org/10.1136/bmjopen-2021-054669>.
 26. Higashiya K, Ford J, Yoon HC. Variation in positivity rates of computed tomography pulmonary angiograms for the evaluation of acute pulmonary embolism among emergency department physicians. *Perm J*. 2022;26(1):58–63. <https://doi.org/10.7812/tpj/21.019>.
 27. Wichmann RM, Fernandes FT, Chiavegatto Filho ADP, et al. Improving the performance of machine learning algorithms for health outcomes predictions in multicentric cohorts. *Sci Rep*. 2023;13(1):1022. <https://doi.org/10.1038/s41598-022-26467-6>.
 28. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. <https://doi.org/10.1186/s12916-019-1426-2>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

