Article published online: 2024-12-10

From Code to Clots: Applying Machine Learning to Clinical Aspects of Venous Thromboembolism Prevention, Diagnosis, and Management

Pavlina Chrysafi[1] Barbara Lam[2,3] Samuel Carton[4] Rushad Patell[2]

[1] Department of Medicine, Mount Auburn Hospital, Harvard Medical School, Cambridge, Massachusetts, United States
[2] Division of Hemostasis and Thrombosis, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States
[3] Division of Clinical Informatics, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States
[4] Department of Computer Science, College of Engineering and Physical Sciences, University of New Hampshire, Durham, New Hampshire, United States

Address for correspondence Rushad Patell, MD, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02215, United States (e-mail: rpatell@bidmc.harvard.edu).

Hamostaseologie 2024;44:429–445.

Abstract

The high incidence of venous thromboembolism (VTE) globally and the morbidity and mortality burden associated with the disease make it a pressing issue. Machine learning (ML) can improve VTE prevention, detection, and treatment. The ability of this novel technology to process large amounts of high-dimensional data can help identify new risk factors and better risk stratify patients for thromboprophylaxis. Applications of ML for VTE include systems that interpret medical imaging, assess the severity of the VTE, tailor treatment according to individual patient needs, and identify VTE cases to facilitate surveillance. Generative artificial intelligence may be leveraged to design new molecules such as new anticoagulants, generate synthetic data to expand datasets, and reduce clinical burden by assisting in generating clinical notes. Potential challenges in the applications of these novel technologies include the availability of multidimensional large datasets, prospective studies and clinical trials to ensure safety and efficacy, continuous quality assessment to maintain algorithm accuracy, mitigation of unwanted bias, and regulatory and legal guardrails to protect patients and providers. We propose a practical approach for clinicians to integrate ML into research, from choosing appropriate problems to integrating ML into clinical workflows. ML offers much promise and opportunity for clinicians and researchers in VTE to translate this technology into the clinic and directly benefit the patients.

Keywords

- venous thromboembolism
- machine learning
- artificial intelligence
- prediction modeling

Introduction

Venous thromboembolism (VTE), including deep vein thrombosis (DVT) and pulmonary embolism (PE), is a significant cause of morbidity and mortality around the world.[1] It is the third leading cause of death from vascular disease, after myocardial infarction and stroke.[2] Moreover, it is the most common preventable cause of death among patients who are hospitalized.[3] Even with appropriate recognition and treatment, there can be long-term complications such as anxiety, postthrombotic syndrome, and chronic thromboembolic pulmonary hypertension.[4] The rising global incidence of VTE is driven by risk factors, which include surgery, hospitalization, cancer, estrogen exposure, and obesity.[1] Preventing VTE is relevant for providers of all specialties, and accurate diagnosis and management can improve outcomes.

received
July 15, 2024
accepted after revision
September 14, 2024

Georg Thieme Verlag KG, Oswald-Hesse-Straße 50, 70469 Stuttgart, Germany

DOI https://doi.org/10.1055/a-2415-8408.
ISSN 0720-9355.

Machine learning (ML) refers to the ability of mathematical algorithms to analyze large amounts of data to detect patterns and make predictions. Its use is well established in everyday parts of our lives including map navigation, facial recognition for security, and even autocorrection when communicating via text messaging. This transformative technology is rapidly gaining traction in healthcare as well and moving from computational labs to the bedside.[5,6] There is considerable interest in developing ML systems for VTE management specifically due to the complex and multivariate factors influencing VTE risk, the availability of multidimensional data in electronic health records (EHRs), and the potential for rapid and early diagnostics through ML integration with imaging.[7]

If ML achieves its promise in the field of VTE, clinicians and patients stand to benefit dramatically. This narrative review describes clinical applications of ML in VTE, exploring both current evidence and potential future directions.

A Primer on Machine Learning Terminology for the Clinician

ML is a subfield of artificial intelligence (AI), which is broadly defined as the capability of a computer to imitate intelligent human behavior. ML specifically describes the use of advanced statistical models to identify patterns in large amounts of data. Although the field of ML has been present for decades, it has gained increasing attention in recent years as computing power has become more affordable, allowing organizations to develop larger and more capable models.[8]

One major division in ML is between classical models and neural net-based models, commonly termed "deep learning"[9,10] (►**Supplementary Table S1**, available in the online version only). Classical models include linear models, such as linear regression and logistic regression, and nonlinear models, such as decision trees, random forests, and k-nearest neighbors. Linear models have long been used to design predictive clinical scores such as the Wells and Geneva scores for VTE.[11,12] These methods rely on the manual selection of variables that are then given specific weights, allowing clinicians to readily interpret the inner workings of the model. More complex nonlinear models are more capable of identifying variables and their relationships within the data, often leading to less interpretable but improved predictive performance. Classical nonlinear models such as random forests and gradient boosting exhibit greater ability in handling complex data structures, though they still depend on manually designed features.[13,14] Deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers[15] are complex, large, and radically nonlinear, and represent the state-of-the-art in most areas of ML (although tree-based methods such as random forests are still competitive for tabular tasks such as predicting VTE risk from numeric features).[16] Classical models are generally available in well-established, easy-to-use toolkits such as scikit-learn,[17] while deep learning models require the use of more engineering-intensive libraries such as PyTorch[18] or TensorFlow.[19]

For the purpose of this review, we have chosen to exclude classical linear models due to their simplicity and limited ability to learn complex patterns from data, focusing instead on classical nonlinear and deep learning models. We also discuss the newest paradigm of ML, known as generative AI, and its potential impact on the field of VTE. Generative AI involves massive general-purpose pretraining to deep learning models, to produce models capable of performing a variety of different tasks without any additional training.[20]

Another key division in ML is the training regime: supervised learning versus unsupervised learning.[21,22] In supervised learning, the model is trained on a labeled dataset. In other words, each example in the dataset is associated with a labeled outcome. In the field of healthcare, these labels are often physician-adjudicated and considered to be the gold standard. These datasets can then be used to train algorithms in tasks such as classification or regression. In unsupervised learning, there are no given labels, and the model is allowed to discover patterns in the data. These models use algorithms such as clustering, association, and dimensionality reduction. Semi-supervised learning is an approach that combines both labeled and unlabeled data, leveraging the model's ability to identify new patterns in the data while reducing the burden of labeling every example.[23] In deep learning, models are often subjected to unsupervised pretraining on a large unlabeled corpus before the given supervised fine-tuning on a smaller labeled dataset.[24,25] In generative AI, these divisions begin to blur even further, with models implicitly learning from massive unsupervised pretraining to perform tasks that traditionally would have required supervised fine-tuning.[26]

Classical ML models rely on feature selection and engineering, which aim to identify the smallest number of relevant variables for the model.[27] There are several different methods of feature selection, ranging from manual selection based on domain expertise to automated methods based on statistical association and collinearity such as mutual information and forward feature selection.[22] Neural net-based deep learning methods have de-emphasized the notion of feature engineering in favor of "representation learning," where the model implicitly learns useful intermediate representations of the data without explicit feature manipulation. It should be noted, however, that the de-emphasis of feature engineering in deep learning is partially a function of the types of data on which deep learning models tend to excel (e.g., text and images rather than tabular data), where explicit feature engineering would be impractical anyway.

Finally, ML models are typically first trained on a portion of the dataset (training set) and then evaluated on a separate portion of the dataset iteratively as parameters are adjusted (validation set).[28] The final dataset that is used to test the model's unbiased accuracy should only include data the model has not previously seen (test set). The metrics used to assess performance depend on the dataset used and the specific clinical objective of the model. Accuracy, precision, recall (a.k.a. sensitivity), specificity, area under the curve (AUC), and F1 (a balance of precision and recall) are performance metrics used frequently.[28]

Current Machine Learning Approaches to Venous Thromboembolism

The ability of ML to process large amounts of high-dimensional data can help identify new risk factors and better risk stratify patients for thromboprophylaxis, interpret medical imaging, and contribute to the shift toward personalized medicine by tailoring treatment according to individual patient needs. Moreover, ML-based natural language processing (NLP) enhances the ability to identify VTE events from clinical and radiological notes that can help build efficient and timely surveillance systems, imperative for epidemiological studies and clinical research.

Risk Prediction for VTE

Thromboprophylaxis is recommended for hospitalized patients at moderate to high risk for VTE and can take various forms, including pharmacologic or mechanical measures.[29] Risk factors such as older age, higher weight, and critical illness can put patients at increased risk, and guidelines recommend an individualized approach to risk stratification.[29,30] A significant challenge lies in balancing the risk of thrombosis against the potential risk of bleeding, particularly in patients with cancer or recent surgery who are susceptible to complications.

At the time of this review, there are more than 35 studies in the literature that leverage ML to predict VTE risk. Models have been developed in various patient populations including postoperative patients,[31,32] hospitalized inpatients,[33–36] and patients with active malignancy.[37,38] A frequently used approach for risk stratification is the random forest model.[32–35,37,38] Key features include patient demographics, laboratory tests, comorbidities, and group-specific information such as the mechanism of trauma in orthopedic patients or tumor characteristics in patients with cancer (►Table 1).

For surgical patients, models have been developed to evaluate the risk of VTE after various procedures including ankle fracture surgery, knee replacement, and hip replacement[31,39] (►Table 1). Ding et al[31] found that XGBoost (a tree-based method) outperformed the traditional Geneva score in predicting VTE after hip arthroplasty, with an AUC of 0.982 compared to 0.552 on their dataset (►Table 1). In non-orthopedic patients, models have been developed for patients undergoing weight loss surgery, inguinal hernia repair, and neurosurgery[31,32,40–42] (►Table 1). Across these studies, models have shown an AUC ranging from 0.65 to 0.989. XGBoost performed best with an AUC of 0.989 in predicting VTE after radical gastrectomy[32] (►Table 1).

In patients with cancer, most models have focused on specific cancer types such as lung, colorectal, gastric, or ovarian cancer. A meta-analysis showed that random forest and logistic regression achieved the highest AUC (≥0.90) in detecting VTE in patients with lung cancer.[43] The aforementioned XGBoost exhibited an AUC of 0.990 in predicting postoperative VTE in patients with colorectal cancer, compared to an AUC of 0.646 for the Khorana and an AUC of 0.769 for the Caprini scores[37] (►Table 1). In an attempt to create a universal model for all cancer types, Mantha et al developed a deep learning model to predict VTE in individuals with solid malignancies[38] (►Table 1). The model incorporated demographic data, cancer-specific information including next-generation sequencing data, and laboratory results. In a validation study with 5,951 individuals, the model achieved a concordance index of 0.72 and outperformed the Khorana score in two other external cohorts.

Thromboprophylaxis during hospitalization is a priority and often linked to quality measures as well as reimbursement strategies.[44,45] Thus, researchers have sought to utilize ML approaches to improve VTE prediction during acute illness by combining EHR data with traditional logistic-based risk assessment scores (such as CAPRINI).[33] ML approaches have been demonstrated to perform superiorly to such risk assessment models in this population.[34] Park et al developed a random forest model derived from EHR data to identify patients at risk for 30-day VTE and readmission after discharge, helping to select patients who may benefit from extended thromboprophylaxis[36] (►Table 1).

A meta-analysis of 20 studies across various patient populations that use ML models for VTE prediction found that ML models that use more modern approaches demonstrated a higher pooled AUC of 0.79 compared to models that use the traditional method of logistic regression.[46] A majority of studies had not been externally validated, however, raising concerns about the generalizability of these findings in clinical settings. Even prior to the rise of modern ML approaches, there were numerous VTE models and a lack of clarity into which models should be used and when.[47] The proliferation of ML models that target select subpopulations only adds to this challenge. Overall, the models described above show excellent performance for VTE prediction when tested internally, outperforming traditional risk prediction scores, but their performance tends to drop in external validation cohorts. Another challenge is that these models are usually designed for specific populations. This specificity poses a problem when, for example, a patient arrives at the emergency department with general symptoms (such as dyspnea or leg swelling) and does not fit neatly into categories like a patient with lung cancer or a patient after orthopedic surgery. Therefore, we need models that can be applied to a wide range of populations to ensure they can be effectively integrated into clinical settings. Chen et al developed a gradient-boosting model predicting VTE in a diverse patient population[48] (►Table 1). Their model achieved AUC values between 0.8 and 0.83 during internal validation, compared to 0.63 for the PADUA score. In external validation, their model achieved an AUC of 0.72 to 0.82 compared to an AUC of 0.61 to 0.77 for the PADUA score. Future studies should focus on representing larger populations and external validation on multiple datasets because universal models could confirm similar approaches can be applied to other settings.

Diagnosis of VTE

Suspicion for VTE typically arises from the clinical presentation and the patient's risk factors. Guidelines recommend calculating a probability assessment using validated scoring

Table 1 Overview of selected studies evaluating ML models in different VTE applications

Study	Study design	Population	Model objective	Dataset size	Data type	ML model	Performance	Comparison with clinical score or clinicians[a]	Externally validated?
Risk prediction for VTE									
Ding et al[31] PMID: 37394825	Retrospective, single center	Orthopedic patients undergoing hip arthroplasty	Predict VTE in patients undergoing hip arthroplasty	Training: 1,037 Validation: 444	Clinical features from EMR	Best performing: XGBoost Other models used: LR, MLP, AdaBoost, GBC, KNN	AUC: 0.98 Sensitivity: 0.91 Specificity: 0.99	Geneva Rapt Plos	No
Liu et al[32] PMID: 37228741	Retrospective, multicenter	Patients undergoing radical gastrectomy	Predict VTE in patients undergoing radical gastrectomy	Internal Training: 873 Validation: 366 External: 1,239	Clinical features from EMR	Best performing: XGBoost Other models used: RF, SVM, KNN	AUC: 0.91	NA	Yes
Qin et al[37] PMID: 36915444	Retrospective, single center	Patients with colorectal cancer postoperatively	Predict VTE in patients with colorectal cancer postoperatively	Training: 833 Validation: 358	Clinical features from EMR	Best performing: XGBoost Other models used: RF, SVM, LR, MLP, LSTM	AUC: 0.91 Sensitivity: 0.77 Specificity: 0.93	Caprini Khorana	No
Mantha et al[38] PMID: 37214902	Retrospective, single center	Patients with cancer	Predict VTE in patients with cancer	Internal Training: 23,800 Validation:5,951 External External MSK: 6,249 ONCOTHROMB: 358	Clinical and genetic features from EMR	Best performing: Deep-Hit Other models used: RF, Fine-Gray Regression	0.72 (reported as C-index)	Khorana	Yes
He et al[33] PMID: 33971809	Retrospective, single center	Trauma patients	Predict VTE in hospitalized trauma patients	Internal Training: 632 Validation: 271	Clinical and genetic features from EMR	Best performing: Caprini + Lasso + RF Other models used: RF combined with various feature selection methods (Lasso, Ridge, ElasticNet, LR, Mutual Information Entropy)	AUC: 0.79 PPV: 0.75 NPV: 0.29	Caprini score with various ML model combinations	No
Nafee et al[34] PMID: 32110753	Retrospective, multicenter	Hospitalized acutely ill medical patients	Predict VTE in acutely ill medical patients	Internal dataset size: 6,454 (unclear training and validation proportion)	Clinical and genetic features from EMR	Ensemble learning models combining estimates from generalized additive models, Elastic Net, Extreme Gradient Boosting, RF, Bayesian LR, Simple Classification Tree	AUC: 0.69	IMPROVE	No
Sheng et al[35] PMID: 37705687	Retrospective, single center	Patients being admitted to the hospital	Predict VTE at admission	Internal Training: 2,154 Validation: 924	Clinical and genetic features from EMR	Best performing: RF Other models used: LR, XGB	AUC: 0.80 Sensitivity: 0.81 Specificity: 0.68 PPV: 0.19	Caprini	No
Park et al[36] PMID: 33617689	Retrospective, single center	Patients being discharged from the hospital	Predict 30-day readmission with VTE	Internal Training: 111,163 Validation: 47,641	Clinical and genetic features from EMR	Best performing: BRF Other models used: LR, MLP	AUC: 0.84 Sensitivity: 0.74 Specificity: 0.80	NA	No
Chen et al[48] PMID: 38095106	Retrospective, single center	Individuals from 3 biobanks (MSDW, UK Biobank, and All of Us)	Predict diagnosis and 1-y risk of VTE in a diverse patient population	Internal Training: 143,101 Validation: 15,900 External UK Biobank: 401,723	Clinical and genetic features from EMR	LightGBM	AUC: 0.83 for diagnosis Sensitivity: 0.95% Specificity: 0.39% PPV: 0.11 NPV: 0.99	Modified Padua	Yes
Diagnosis of VTE									
Schmuelling et al[54] PMID: 34157638	Retrospective, single center	Patients who had a CTPA in the ED between 4/2018 and 6/2020	Detect PE on CTPA (and evaluate effect on report time, time to AC and patient turnaround times in the ED)	1,808 CTPA Unclear distribution between training and validation	CTPA	Deep Learning (not further specified)	Sensitivity: 0.79 Specificity: 0.95 PPV: 0.82 NPV: 0.94	Radiologists	No

Table 1 (*Continued*)

Study	Study design	Population	Model objective	Dataset size	Data type	ML model	Performance	Comparison with clinical score or clinicians[a]	Externally validated?
Jamin et al[56] PMID: 37370233	Retrospective, single center	Patients with lower-limb DVT	To evaluate the effectiveness of bi-dimensional entropy measures in detecting PE using US images of DVTs	32 patients (16 with PE and 16 without) Total 1,490 US images. Unclear distribution between training and validation	US	Bi-dimensional entropy measures Best-performing: FuzEn2D Other: DispEn2D	AUC: 0.72	NA	No
Fisher et al[55] PMID: 8628857	Retrospective, single center	Patients who underwent V/Q scan between 1991 and 1995	Detect PE on V/Q scans	Internal Training: 150 V/Q scans Testing: 30 V/Q scans	V/Q scans	Artificial neural network	Correlation coefficient of 0.77	No	No
Somani et al[92] PMID: 35355847	Retrospective, multicenter	Patients with moderate-high suspicion for PE that underwent CTPA and at least one EKG	To evaluate the role of a fusion multimodal model combining EKG and EHR to detect PE	Internal Training: 19,065 patients Testing: 2,118	ECGs and data from EHR	CNN for EKG component XGBoost for the EHR component	AUROC: 0.84	Well, Geneva, PERC, 4PEPS	No
VTE prognosis									
Deng et al[64] PMID: 36272528	Retrospective, single center	Allogeneic stem cell transplant patients with VTE	Identify post-transplantation VTE patients at high risk of death to improve survival through early intervention	Training: 205 patients Validation:31 patients	Clinical and laboratory features from EMR	Best performing: LR classifier Other models: XGBoost, LightGBM, RF, Adaptive Boosting, Naïve Bayes, Multilayer Perceptron, K-Neighbors classifier	AUC: 0.87 Sensitivity: 0.94 Specificity: 0.78 PPV: 0.66 NPV: 0.88	NA	No
Mora et al[65] PMID: 34107539	Retrospective, multicenter	Patients with acute PE and premature anticoagulation discontinuation (<90 d)	Predict the composite outcome of fatal PE or recurrent VTE in 30 d after premature discontinuation of anticoagulation	Training: 944 patients Validation: 404 patients	Clinical features from EMR	Best performing: NN Other models: RUSBoost, SVM, Decision Tree, k-Nearest Neighbors Algorithm	AUC: 0.96 Sensitivity: 1.00 Specificity: 0.94 PPV: 0.33-0.79 NPV: 1.00	NA	No
Cahan et al[123] PMID: 37160926	Retrospective, single center	Patients who underwent CTPA between 2012 and 2018	Develop and evaluate a fusion multimodal model for classification of PE	Training: 242 CTPAs Validation: 44 Testing: 72	CTPAs and EHR	Fusion of 3D CNN and TabNet Fusion models, including TabNet, CNNs, XGBoost, and SANet.	AUC: 0.96 Sensitivity:0.90 Specificity: 0.94	PESI and sPESI	No
Risk for treatment complications									
Mora et al[67] PMID: 36942630	Retrospective, multi-center	Patients with acute VTE	Detect patients with VTE and increased risk for major bleeding during the first 3 mo of anticoagulation	Internal Training: 34,7111 Validation: 14,876 External COMMAND-VTE database: 401,723	Clinical and laboratory features from EMR	Best performing: XGBoost Other models: SVM, NN, Decision Tree, KNN	AUC: 0.91 Sensitivity: 0.33 Specificity: 0.93 PPV: 0.10 NPV: 0.98	RIETE VTE-BLEED	Yes
Grdinic et al[69] PMID: 38184201	Retrospective, single center	Patients with cancer-associated thrombosis	Predict bleeding in patients with CAT receiving anticoagulation at 90, 365, and 455 d	Training: 756 patients Validation: 324 patients	Clinical and laboratory features from EMR	Best performing: XGBoost Other models: Lasso LR, Ridge LR, RF	0.64 at 90 days Sensitivity: 0.17 PPV: 0.18	CAT-BLEED	No
Fard et al[70] PMID: 38642704	Retrospective, single center	Patients with weakly provoked/unprovoked	To predict the risk of major bleeding in	Training: 1,779 patients		Best performing: Ensemble of feedforward and recurrent	AUC: 0.82 Sensitivity: 0.61	RIETE VTE-BLEED	No

(*Continued*)

Table 1 (*Continued*)

Study	Study design	Population	Model objective	Dataset size	Data type	ML model	Performance	Comparison with clinical score or clinicians[a]	Externally validated?
		VTE who required extended anticoagulant therapy	patients on extended anticoagulation therapy by incorporating patient time series follow-up data	Validation: 763 patients	Clinical, laboratory, and genetic features from EMR	neural network Other models: Baseline-ANN, LastFUP-ANN, FUP-ANN	Specificity: 0.82 PPV: 0.13		
Anticoagulation choice and duration									
Choi et al[71] PMID: 38105280	Retrospective, single center	Patients who were prescribed warfarin (not selectively in patients with VTE)	To determine warfarin discharge dose duration to prevent adverse effects	Internal Training: 2,534 patients Validation: 634 patients External MIMIC-III Dataset 981 patients	Clinical features from EMR	Best performing: XGBoost and ANN Other models: RF, LR	NA	Expert physicians	Yes
Detecting VTE in EMR									
Chapman et al[124] PMID: 21459155	Retrospective, single center	Patients who underwent CTPA for PE suspicion	Develop a CTPA report classifier for identifying PE and distinguishing between acute and chronic	Training: 250 reports Testing: 650 reports	CTPA reports	Best performing: PEFinder Others: Unigram and Bigram Naive Bayed	Sensitivity: 0.98 Specificity: 0.89 PPV:0.83	NA	No
Banerjee et al[125] PMID: 29175548	Retrospective, single center	Patients with CTPA reports	Develop a model to detect PE in CTPA reports, the difference between acute and chronic, and the location of PE	Internal (Stanford database) Training: 3,512 reports Testing: 1,000 reports External: 858 reports (UPMC)	CTPA reports	Best performing: IWE model Other models: rule-based PEFinder model and out-of-the-box word2vec model	PE acute: AUC 0.96 PE present: AUC 0.95 PE subsegmental: AUC 0.92	NA	Yes
Chen et al[126] PMID: 29135365	Retrospective, single center	Patients with CTPA reports between 1998 and 2016	Evaluate a CNN model (compared with a traditional NLP model) in extracting PE findings from CTPAs	Internal Training (Stanford): 2,500 reports External (UPMC): 859 reports	CTPA reports	Best performing: CNN model Other models: rule-based PEFinder model	AUC: 0.97 Sensitivity: 95% Specificity: 99.7%	NA	Yes

Abbreviations: AdaBoost, adaptive boosting; ANN, artificial neural networks; AUC, area under the curve; BP-Network, Backpropagation Network; BRF, balanced random forest; Caprini + Lasso + RF, Caprini score combined with Lasso regression and random forest; CNNs, convolutional neural networks; CT, classification tree; Deep-Hit, deep learning model for survival analysis (specific model name); DispEn2D, dispersion entropy in two dimensions; DVT, deep vein thrombosis; ElasticNet, elastic net regression; EMR, electronic medical records; FuzEn2D, fuzzy entropy in two dimensions; GBC, gradient boosting classifier; IWE, intelligence word embedding; KNN, k-nearest neighbors; LDA, linear discriminant analysis; LightGBM, light gradient boosting machine; LR, logistic regression; LSTM, long short-term memory; ML, machine learning; MLP, multilayer perceptron; naive Bayes, naive Bayes classifier; NN, neural networks; NPV, negative predictive value; PE, pulmonary embolism; PPV, positive predictive value; RF, random forest; RO, recurrent optimization; RUSBoost, random under-sampling boost; SVM, support vector machine; VTE, venous thromboembolism; XGBoost, extreme gradient boosting.
Notes: This selection was based on a PubMed search for "machine learning *and* venous thromboembolism" which yielded 118 studies. Titles were screened to identify those evaluating ML models for VTE applications. Studies were included based on relevance, excluding those with similar populations or published in lower-impact journals. Performance metrics describe the best-performing model on the internal dataset.
[a]In all cases, best-performing models outperformed clinical scores.

systems such as the Wells score.[49,50] Biomarkers such as D-dimer can further risk stratify which patients need further testing, although diagnoses are confirmed through imaging studies: duplex ultrasonography for DVT and computed tomography pulmonary angiography (CTPA) for PE. Approaches for integrating ML into improving diagnostic accuracy range from assigning diagnostic probabilities at presentation,[51] development of biomarkers,[52] to ML-based approaches to optimize radiologic diagnosis of VTE[53–56] (►**Table 1**).

Most ML approaches in this area have focused on developing models that can be paired with a radiologist to help expedite reads (►**Table 1**). Nakayama et al leveraged a ResNet101 model to classify ultrasound images positive for DVT among disaster victims.[57] Meta-analyses to evaluate the effectiveness of a pooled deep learning model in detecting PE included 36,847 CTPA found that the model had a sensitivity of 88% and a specificity of 86%,[58] compared to the performance of radiologists, whose sensitivity is estimated to be 67 and 87% and specificity 89 to 99%.[58–61] Notably, ML models to aid PE diagnosis in the emergency department demonstrated excellent model performance but did not lead to statistically significant reductions in study reading time, report communication time, or time to anticoagulation, highlighting the gap between technical and clinical success[54] (►**Table 1**). Future studies should focus on testing in real-world clinical environments so that we can begin to close this gap.

Exploring how we can use ML models for difficult-to-detect findings may also improve impact. For example, bi-dimensional entropy measures in ultrasound images are being explored to detect which DVTs were likely to be associated with PE,[56] an approach that could improve patient outcomes, decrease healthcare costs, and minimize radiation exposure by avoiding low-yield scans (►**Table 1**). There is yet a lack of models for detecting chronic PE and CTEPH that have been identified as potential foci for future studies.[62]

Management of VTE

The mainstay of VTE treatment is anticoagulation, aimed at preventing clot extension and the formation of new clots.[49] Determining the duration of anticoagulation treatment depends on the patient's risk of recurrence and bleeding.[49] However, risk factors are often dynamic and can change over time as they are influenced by age, new comorbidities, and changes in medications which can make individualized decisions imperative but also challenging.

ML has been used to predict VTE severity, which can guide treatment. In certain high-risk cases of PE, more aggressive interventions such as thrombolysis or thrombectomy may be needed.[63] Deng et al developed a model forecasting 2-year overall survival after VTE in patients who received allogeneic hematopoietic stem cell transplantation (HSCT), stratifying patients into low-, intermediate-, and high-risk clusters[64] (►**Table 1**). Such prognostic tools could help clinicians start treatment promptly for higher-risk patients and adjust the risk–benefit discussion regarding interventional procedures or anticoagulation in patients with thrombocytopenia undergoing HSCT. A frequent and relevant clinical question pertains to the duration of anticoagulation therapy, and when it is safe to discontinue these medications. Neural network models have been used to identify a composite outcome of VTE-related death or VTE recurrence in patients with VTE who prematurely discontinued anticoagulation[65] (►**Table 1**). This model outperformed traditional logistic regression with an ROC curve of 0.96 compared to 0.76.

Determining a patient's risk of bleeding can be challenging to quantify, and existing risk prediction tools such as HAS-BLED and VTE-BLEED demonstrate modest performance.[66] Mora et al developed an XGBoost model capable of predicting major bleeding events in patients receiving anticoagulation for VTE, initially outperforming VTE-BLEED on an internal dataset but with comparable performance on an external dataset[67] (►**Table 1**). Cancer patients are at increased risk for thrombosis but also bleeding.[68] The conventional CAT-BLEED score was compared with several different ML modeling approaches to bleeding events up to 455 days after an episode of cancer-associated thrombosis[69] (►**Table 1**). All models outperformed CAT-BLEED, but they used over 400 variables including clinical data, biochemistry, and diagnosis codes, which make the model challenging to replicate in other settings. Recognizing that the risk of bleeding can change over time, Shahryari Fard et al developed a feed-forward and RNN model that used data from multiple time points to predict major bleeding in patients on extended anticoagulation, achieving an AUC >0.80[70] (►**Table 1**). Thus, this could be envisioned as an approach that dynamically pulls data at every visit and continuously updates over time to enhance the prediction accuracy.

There are crucial therapeutic choices related to agents and dose or route of administration for patients who are treated with anticoagulation that could be optimized by integrating ML. In one approach, ML models have been developed to predict the starting warfarin dose based on clinical data and were shown to outperform physician estimates[71] (►**Table 1**). Ravvaz et al created a Bayesian network model leveraging pharmacologic and genomic data to tailor optimal warfarin dosing to different subpopulations and were able to demonstrate improved time in the therapeutic range.[72] A random forest model used a causal inference approach to determine whether patients would most benefit from apixaban versus rivaroxaban treatment in reducing the risk of stroke and major bleeding.[73] Although these studies focused on patients with atrial fibrillation, they underscore the potential for ML to personalize treatment choices and the need for further studies in VTE specifically exploring not only anticoagulation but also which patients might benefit from intervention.

Identifying VTE in the EHR

Identifying VTE cases in the EHR is fundamental for disease monitoring to ensure prevention measures and working. An accurate representation of cases also facilitates further clinical research, identifying datasets that can be used for external validation of ML models and the development of new models. Traditional methods for identifying VTE cases, such as manual chart review, are time-consuming. Using diagnosis codes can speed the process but has been shown to have poor predictive value (PPV).[74,75] NLP, a type of AI that focuses on using computers to extract and interpret text, has emerged as a promising solution for identifying information in the EHR including cases of VTE.

Early forms of NLP were "rule-based" and relied on a series of *if, then* rules to parse text. These lists of rules required provider expertise to create, had to be customized to the preferred language of the institution, and could become long and unwieldy. PEFinder and MedLEE were two well-known examples of rule-based NLP to detect VTE that performed very well but were time-intensive to create and transport to other settings.[76] As more advanced statistical ML methods have developed, researchers have explored using other NLP techniques to identify VTE cases (►Table 1). A meta-analysis of eight studies demonstrated a pooled sensitivity of 0.931, specificity of 0.984, and PPV of 0.910.[77] The most effective models used a newer text preprocessing method known as vectorization, where words are represented as mathematical vectors, and newer deep learning approaches including convolutional neural networks. Despite the excellent pooled performance, the review noted significant heterogeneity across studies and emphasized the need for publication standards to meaningfully advance the field. Of note in a study specifically looking at patients with cancer, the combined use of diagnosis billing codes and NLP to create computable phenotypes performed better than either approach individually, highlighting that this strategy might be optimal.[78] The use of NLP to identify bleeding events is also an important step as these are particularly challenging to capture using billing codes alone, given the heterogeneity compared to thrombotic events. Although studies are limited, there is also promise for this approach to identify hemorrhagic events from clinical unstructured data.[79,80]

Future Directions in Machine Learning for VTE

The dominant trend in contemporary ML has been the rise of pretrained generative models, which are trained on large-scale unlabeled data to capture (and be able to generate from) the underlying data distribution. The most visible example of this approach is large language models (LLMs) such as ChatGPT[81] and Gemini,[82] sometimes also known as foundation models,[83] which are trained on large-scale corpora of text to become robust text-comprehension agents. However, generative approaches can also be applied to other data types for tasks such as designing new molecules (including new anticoagulants), creating digital twins for synthetic clinical trials, and expanding existing datasets (►Fig. 1).

Large Language Models and VTE

The large-scale pretraining of LLMs allows them to act as robust linguistic agents with only written prompts. This learning paradigm is termed "in-context learning," where "few-shot" and "zero-shot" prompting refer to the inclusion versus exclusion of minimal training examples ("exemplars") in the inference-time input to the model. The process of optimizing written prompts is known as "prompt engineering." These model's performances can vary widely depending on their inputs, and some have argued that prompt engineering is quickly becoming a core skill for both clinicians and researchers.[84] In general, the more direction the model is provided upfront, the

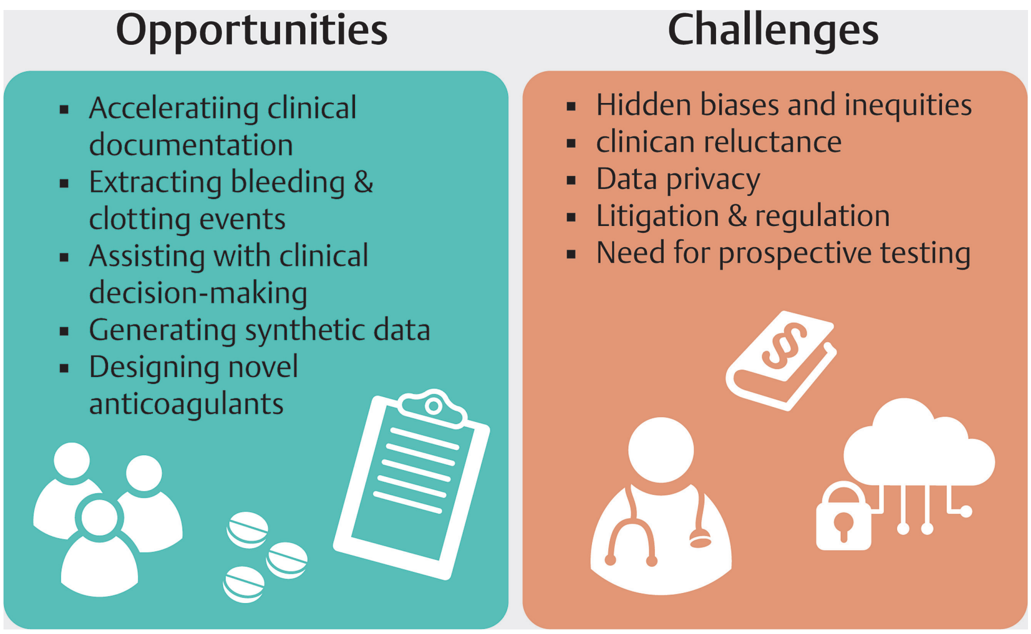

Fig. 1 Future opportunities and challenges in machine learning for venous thromboembolism.

more likely it is to perform well and minimize hallucinations.[85]

However, the largest models of these kinds, such as ChatGPT and Gemini (100+ billion parameters), are proprietary in nature and too large in any case to run on the hardware available to most academic or medical teams. They are typically accessed via application programming interfaces (APIs) with cost based on use, which can quickly become expensive (►Fig. 1). These interfaces also involve sending data to an external server, raising privacy risks for health-protected data (►Fig. 1). Smaller LLMs (1–30 billion parameters) such as LlaMA[86] and OLMo[87] are available open-access for download. Although smaller models are less capable of in-context learning compared to larger ones, they can be customized and used on private servers and, with task-specific training, can often outperform larger, more general-purpose models[88] (►Fig. 1).

Pretrained LLMs excel at generic tasks within the ambit of their pretraining data. So, an LLM would not have a firm basis for predicting risk scores for VTE from numeric features but could be expected to parse and generate written clinical notes (►Fig. 1). Van Veen et al, for example, showed that GPT-3.5 compared favorably to human experts at summarizing clinical text,[89,90] while Ge et al showed that GPT-4 was able to automatically extract key information from radiology reports.[91] In the context of VTE, an LLM could potentially be used to identify bleeding and clotting events in the medical record with minimal to no training, programming, or data preprocessing (►Fig. 1).

LLMs are becoming increasingly multimodal, with major implications for medicine generally and VTE specifically. Saab et al recently introduced Med-Gemini, a medicine-specialized version of the Gemini model, demonstrating its zero- and few-shot performance across a variety of multimodal tasks including surgical video understanding and radiology image captioning.[82] This trend toward multimodality extends beyond LLMs, with Somani et al using fused representations of both electrocardiogram images and other types of clinical data to diagnose PE.[92] In VTE specifically, this trend carries the potential for zero- and few-shot reasoning over both VTE-related text and images, among other data modalities (►Fig. 1).

Generative AI Outside Foundation Models

Pretrained generative approaches can be applied anywhere there exists a large corpus of unlabeled data that would be helpful for the model to "learn" in a general way before adapting it for a particular purpose. By learning the underlying distribution of this data, generative models can produce useful feature representations of data items—for example, a neural autoencoder was run across a corpus of Simplified Molecular Input Line Entry System (SMILES, a notation for representing molecular structures in text format) representations of chemical compounds to suggest new candidates for anticoagulants[93] (►Fig. 1). Another illustrative use can be to simulate data items for the purpose of counterfactuals to generate "digital twins"[94] (►Fig. 1). Synthetic data generated can be used to produce synthetic data for the purpose of training downstream models. Ktena et al showed that augmenting a training set with images produced by a generative model pretrained on large-scale imaging data improves the fairness of a diagnostic model under distribution shift.[95] All of these uses for non-foundation generative AI have potential applications in the modeling of VTE.

Hacking the Challenges: Hurdles and Solutions for ML VTE Applications

Although there is much promise for ML to be applied directly for clinical applications, the stakes involved in healthcare necessitate early recognition of challenges that need to be addressed for the safe and sustainable development of the field[96] (►Fig. 1).

Availability of High-Quality Datasets

The development and validation of ML models depend on reliable and representative datasets. EHRs are rich repositories of multidimensional data, but data are often stored in unstructured fields, of variable quality, not broadly representative of different patient populations, and require human labeling that can be inefficient and expensive. Building large datasets from multiple institutions is essential, but there are numerous challenges including reconciling different test assays and how clinical variables are interpreted, as well as practical issues such as data storage and how to manage updates over time.[97] Current studies are almost exclusively trained and tested on institution-specific datasets (►Table 1), which makes these models of limited utility outside the environment they were developed.

Moreover, because ML approaches such as deep learning can be prone to overfitting, one dataset is not enough and ideally, there are multiple sources for validation.[98] Health care data are subject to privacy regulations that make transferring large volumes of patient information difficult and complete anonymization of data prior to sharing is inefficient.[99] Using artificially generated data is one solution, and another is to leverage federated learning, a decentralized approach where models are sent to different institutions for training without any exchange of patient data.[100,101]

Validation and Prospective Testing

Clinical ML models can be developed on retrospective data but need to be tested prospectively to ensure they perform reliably in real-world settings.[102] This step is essential to ensure safety, generalizability, and clinically meaningful outcomes. There are several trials registered for ML applications in VTE specifically (►Table 2), and clinicians should evaluate the rigor with which a model was evaluated before using it, just as they would for any other clinical intervention. After a model enters the clinical space, continuous quality assurance and performance monitoring are vital for maintaining effectiveness, safety, and trustworthiness.[103] This monitoring can help detect and correct "model drift" over

Table 2 Ongoing clinical trials in machine learning and venous thromboembolism

Registration number	Population	Primary objective	Country	Study start, status	Status
NCT06341231	Patients with PE and age ≥75, renal insufficiency, or coexisting malignancy	ML models to predict major bleeding, CRNMB, and VTE recurrence from enrollment to 12 mo after anticoagulant therapy	China	2024, recruiting	
NCT05905874	Patients with COPD	ML models to predict DVT and PE in patients with COPD	China	2023, recruiting	Recruiting
NCT06183944	Patients with acute PE	Assess the validity and transferability of previously developed NLP tools for data acquisition and structure from medical reports Hierarchical clustering methods to identify homogeneous groups of patients and predict 6 mo prognosis	France	2023	Recruiting
NCT06284343	Female patients with gynecologic malignancies scheduled for systemic antineoplastic treatment	Development and validation of a risk prediction model for VTE	Italy	2024	Not yet recruiting
NCT05729464	Patients with cancer	Identify risk factors of cancer-associated venous thrombosis and develop a prediction model to assist clinicians in tailoring anticoagulant therapy	China	2023	Recruiting
NCT05860790	Patients undergoing neurosurgery	Development and validation of a risk prediction model for VTE	China	2023	Recruiting
NCT04768036	Hospitalized patients	Evaluate whether an electronic health record-embedded clinical prediction rule model will increase the proportion of hospitalized patients at risk for VTE	USA	2022	Completed

Abbreviations: COPD, chronic obstructive pulmonary disease; DVT, deep vein thrombosis; PE, pulmonary embolism; VTE, venous thromboembolism.
Note: Last reviewed June 2024.

time that can occur due to shifts in data and the working environment.[104]

Risk of Bias and Expanding Existing Health Disparities

Despite its advantages, ML can harbor bias that can worsen disparities in healthcare outcomes, particularly for underrepresented groups.[105] Given the high stakes for patient care, especially as medical applications for AI grow exponentially, it is essential that these tools be tested rigorously for unwanted bias. Development efforts should include strategies to improve and refine the tools to rectify sources of these biases. Innovative approaches for fairness-aware ML include algorithmic auditing to detect bias and techniques like reweighting and re-sampling of data to ensure diversity.[106,107] These approaches have been shown to develop ML models that demonstrate robust performances for VTE prediction across patient subsets including ethnicity and age.[48] Moreover, clinical decision support tools have been shown to be a possible solution to eliminate racial disparities in the use of VTE prophylaxis, highlighting the potential of these technologies to enhance health equity.[108]

Clinician Reluctance and Explainability/Interpretability

In order to gain a better understanding of the stakeholder's attitudes toward the implementation of thrombosis-risk prediction models, we conducted two national surveys: clinician-directed ($n = 607$) and healthcare informaticians ($n = 101$). The majority of both clinicians (70.1%) and informaticians (56%) believed that ML can be used to manage VTE; however, several potential concerns were raised for this to be realized including transparency.[109] Although this can be challenging, particularly in advanced ML development like deep neural networks, applying "explainability techniques" to identify clear and interpretable outputs and key features that influence them can help improve trust in the algorithms' performance by all stakeholders including health care providers and patients.[110] However, the problem of how to impose explanations on intrinsically nonlinear models and present these explanations to human stakeholders in ways that improve their ability to audit models is an open problem and active research area within ML.[111]

Regulatory and Liability Concerns

It is essential that regulatory agencies keep up with the rapid pace of computational progress to ensure clinical applications for AI are standardized, safe, and ethical, and facilitate innovation and implementation of these technologies. Various guidelines have been published for the standardized reporting of medical applications of AI.[112] The U.S. Food and Drug Administration has outlined 10 guiding principles for the development of AI/ML tools in medicine.[113] Litigation claims related to VTE are not infrequent and malpractice suits are commonly around missed or delays in diagnosis and administration of prophylactic anticoagulation.[114,115] Thus, as AI-based solutions enter patient care it will also be essential for legislation to be developed for potential legal issues and liability implications that can arise from medical errors/malpractice.[53]

Practical Steps to Getting Started with Machine Learning in Clinical Research and Applications

Many clinical researchers are interested in exploring ML approaches for their own research in the field of VTE. We highlight key steps and relevant resources as well as learning points from the current state of the literature (►Fig. 2).

Picking the Right Problem to Solve

The first step in ML research is to identify an appropriate problem that addresses an important clinical or operational need.[116] Many tasks, particularly tabular prediction tasks involving discrete numerical or categorical features, can be effectively solved by simpler approaches. Data availability can determine which problem you choose to pursue (►Table 3). In general, the more challenging the task, the more data are needed. A key part of problem exploration is determining what type of data you need, whether you trust the data and can access it, and if it represents your target population well. It is also valuable to determine upfront if there are enough data from a different source that will allow you to externally validate your model, which is an essential step in creating tools that can be used clinically.

Building a Team

An appropriate team is essential to ensuring success. Key contributors include a member with clinical expertise and a member with data science experience. The clinical expert should be involved in every step of model build and development. They are critical during data extraction and cleaning because they can help make decisions around missing data and identify potential errors or biases in the data before it is incorporated into a model. They can help with feature selection if needed, provide input on how much explainability is preferred, and assist with picking clinically relevant performance metrics.[117] The process of obtaining and preparing a dataset for modeling often takes the most amount of time in ML studies. An appropriate data scientist will have experience with data analysis and developing and testing models, ideally with healthcare data specifically.

Computational Resources

Deep learning models generally require more computing power than is available on a typical personal computer. There are three main ways researchers typically secure computational resources: they may purchase and maintain it themselves, purchase temporary use from their institution's own high-performance computing cluster, or purchase temporary use through a cloud service. Clarifying the best approach with your institution's research office is an important step to ensure the method you choose is compliant with

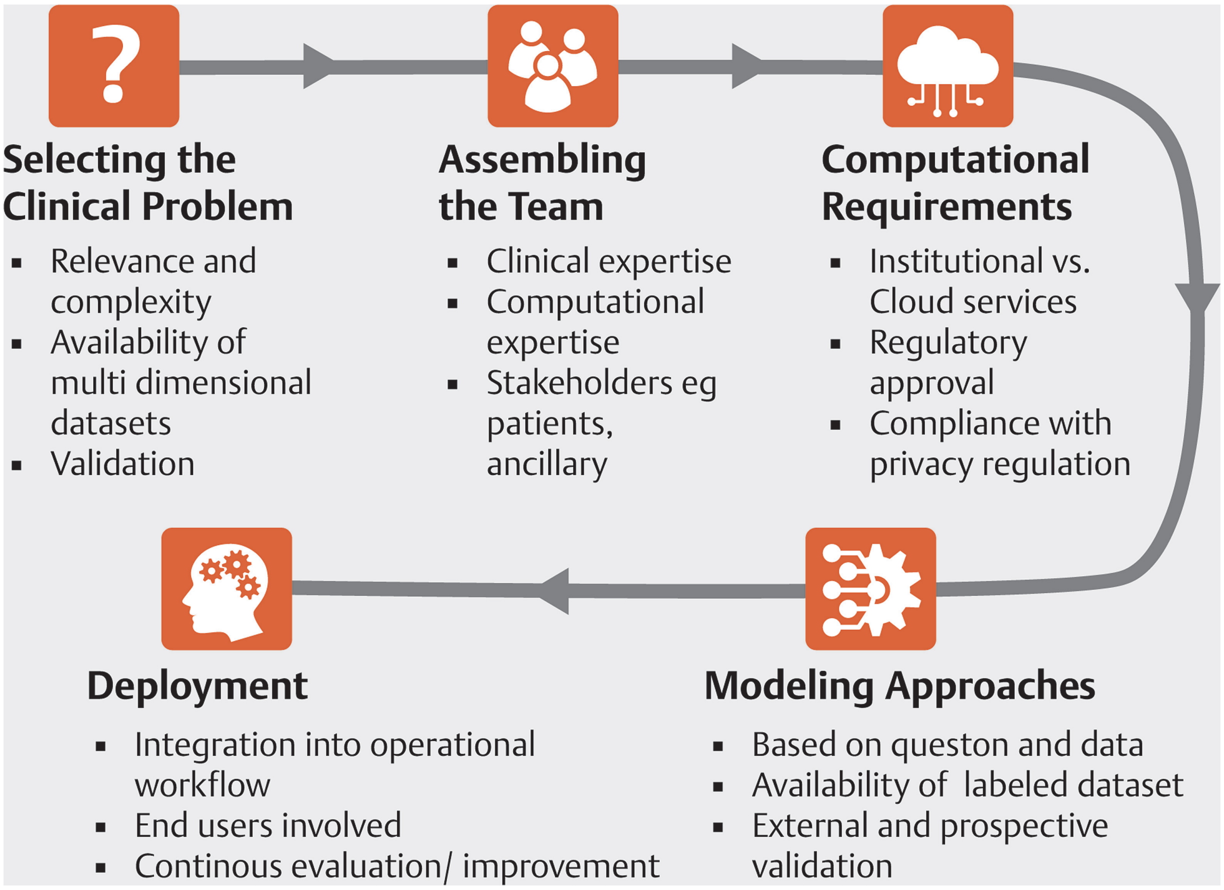

Fig. 2 Practical steps to getting started with machine learning in clinical research and applications.

regulatory needs within your own local and national centers. Of note, some cloud services may include ML tools on their platform that can be used with minimal coding, which can be a useful option.[118]

Modeling Approach

The task, amount and type of data, and available computational resources will all help dictate the approach (►Table 3). Three basic choices include (1) training a classical model such

Table 3 Overview of VTE-related tasks and data modalities

VTE-related task	Commonly used type of data	Examples	Commonly used machine learning approach	Considerations
Risk prediction of VTE	Tabular	Demographics Laboratory data Genetic sequencing	Classical machine learning	This is the most commonly used type of data but is also prone to errors, missingness, and bias. Careful review of data should be done prior to modeling
Case detection in the chart	Text	Clinic notes Radiology reports Discharge summaries	Classical NLP Transformer-based neural network	Rich information about a patient's story can be found in free-text notes. Consideration should be given to what types of notes are included and what information may be excluded. For example, a radiology report describing a "known" pulmonary embolism may not indicate that the event occurred just prior to transfer

Table 3 (*Continued*)

VTE-related task	Commonly used type of data	Examples	Commonly used machine learning approach	Considerations
Diagnosis of VTE	Images	CT images V/Q scans Electrocardiograms	Convolutional neural network	ECGs can be represented as signal amplitudes (1D) over time or as an image (2D) over time. Other imaging modalities, such as CT scans, are 3D in nature. The dimensionality of the image informs the modeling approach
	Video	Ultrasounds Echocardiograms	Convolutional neural network	Video data can present unique challenges. It is inherently multimodal, including both spatial and temporal information, and is often data-intensive
	Time series	Vital signs Laboratory data Electrocardiograms	Recurrent neural network	Granular changes in data can indicate changes in clinical status. A time series approach to representing data may better reflect dynamic clinical decision-making

Abbreviations: ECG, electrocardiogram; NLP, natural language processing; VTE, venous thromboembolism.

as a random forest from scratch, (2) fine-tuning a pretrained deep learning model such as BERT or LlaMA, or (3) using an API-based LLM such as GPT-4. Classical models perform competitively on tabular data such as numeric patient attributes, while neural nets excel at processing text and image data, and can be creatively engineered to, for example, create fused representations of multiple data types.[53] API-based LLMs will excel at generic reading comprehension tasks like extracting structured information from written text and are likely to generalize very well to different datasets. TRIPOD-AI provides guidelines on model development and best practices such as external validation.[119]

Deployment

The development of ML models in healthcare should focus on practicality and the potential for use in real-life settings in order to truly bring the power of ML to patient care. Several clinical trials for testing models are ongoing, and we as a field should continue to push to rigorously test models and explore integrating them into clinical care (►**Table 2**). Deploying model is a broad field and is well covered in other reviews.[120] The team should be expanded to include those who can speak to clinical or operational workflows, help design the interface for the system, and assist with implementation. End users including clinicians and patients can assist with these steps, act as champions, and guide education. As with the implementation of any new technology, understanding and anticipating human behavior is as important as the quality of the technology and should be prioritized. Information technology leaders also need to be involved in discussions around ongoing model quality control and plans for monitoring for bias and performance drift.[121] To address and supervise these aspects, institutions have created AI stewardship committees that can assist with model deployment.[122]

Conclusion

ML has long played a role in the field of VTE, first with logistic regression models powering risk prediction scores and rule-based NLP to identify cases in the chart. As statistical models become more advanced, new opportunities arise for leveraging modern ML approaches for improving clinical care. On the rapidly approaching horizon are generative AI techniques that allow for the creation of new data. Large, existing models can be applied to a variety of tasks without further customization, opening the door for researchers and clinicians outside the field of ML to explore new tools. Ultimately, however, the field of VTE needs to focus next on translating this new technology to the clinic so that providers and patients can see the benefits in their daily lives.

As noted earlier, existing ML approaches to various aspects of VTE are characterized by institutional siloing leading to non-generalizable outcomes and artifacts. Generative AI, in the form of both pretrained foundation models and generative models of other unlabeled VTE-related corpora, promises a way out of this situation. Models within these emerging paradigms have the potential to be generalized across domains. They also have the potential to ease dataset collection by automating chart review, further reducing the problem of data availability. Yet at the same time, researchers seeking to exploit the potential of these models have to grapple with the cost, limited access, and privacy concerns of API-based models, or the technical and computational challenge of fine-tuning smaller foundational models or pretraining novel generative models.

Independent of the emerging capabilities of generative AI, there is a need for researchers to consider the practical applicability of models developed in a research context. When a certain type of model is trained to a certain AUC on a certain VTE-related task on a certain institution's internal data, the barriers preventing physicians at that institution from using that model as a reliable and useful clinical tool must be understood and tackled. Interpretability, fairness, and human factors are all elements of this question that tend to be neglected once the model has been trained and evaluated. Furthermore, when privacy concerns make it impossible to publish the training data and often the model itself, finding ways so that the trained model is useful to physicians at other institutions and its role in pushing the field of ML-powered VTE forward on a broader scale is essential. While generative AI promises potential solutions to these problems, there is an imperative need for future research to realize these solutions.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This study was supported by the Centers for Disease Control and Prevention (CDC), Atlanta, GA, Cooperative Agreement #DD20-2002. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. R.P. is partially funded by the Conquer Cancer Foundation Career Development Award and the National Blood Clot Alliance.

References

1 Lutsey PL, Zakai NA. Epidemiology and prevention of venous thromboembolism. Nat Rev Cardiol 2023;20(04):248–262
2 Sidney S, Lee C, Liu J, Khan SS, Lloyd-Jones DM, Rana JS. Age-adjusted mortality rates and age and risk-associated contributions to change in heart disease and stroke mortality, 2011-2019 and 2019-2020. JAMA Netw Open 2022;5(03):e223872
3 Office of the Surgeon General (US); National Heart, Lung, and Blood Institute (US). 2008
4 Fanikos J, Piazza G, Zayaruzny M, Goldhaber SZ. Long-term complications of medical patients with hospital-acquired venous thromboembolism. Thromb Haemost 2009;102(04):688–693
5 Ferroni P, Zanzotto FM, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast cancer prognosis using a machine learning approach. Cancers (Basel) 2019;11(03):328
6 Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19(01):64
7 Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol 2020;1213:3–21
8 Hirani R, Noruzi K, Khuram H, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. Life (Basel) 2024;14(05):557
9 Guetari R, Ayari H, Sakly H. Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches. Knowl Inf Syst 2023;65(10):1–41
10 Huang S, Arpaci I, Al-Emran M, Kılıçarslan S, Al-Sharafi MA. A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability. Multimedia Tools Appl 2023;82(22):34183–34198
11 Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. Ann Intern Med 2001;135(02):98–107
12 Wicki J, Perrier A, Perneger TV, Bounameaux H, Junod AF. Predicting adverse outcome in patients with acute pulmonary embolism: a risk score. Thromb Haemost 2000;84(04):548–552
13 Pereira FC, Borysov S. Machine Learning Fundamentals. Chapter 2 - Machine Learning Fundamentals. Elsevier; 2019:9–29
14 Schulz MA, Yeo BTT, Vogelstein JT, et al. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. Nat Commun 2020;11(01):4238
15 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv:170603762 [cs] . 2017
16 Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? arXiv;. 2022
17 Scikit-learn. Accessed September 25, 2024 at: https://scikit-learn.org/stable/about.html#citing-scikit-learn
18 Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, Inc. Curran Associates, Inc; 2019:8024–35
19 Abadi M, Agarwal A, Barham P, et al. ensorFlow: Large-scale machine learning on heterogeneous systems. arXiv 2015:1603.04467
20 Salahuddin Z, Woodruff HC, Chatterjee A, Lambin P. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. Comput Biol Med 2022;140:105111
21 Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics 2021;22(04):291–300
22 Tufail S, Riggs H, Tariq M, Sarwat AI. Advancements and challenges in machine learning: a comprehensive review of models, libraries, applications, and algorithms. Electronics (Basel) 2023; 12(08):. Doi: 10.3390/electronics12081789
23 Eckardt JN, Bornhäuser M, Wendt K, Middeke JM. Semi-supervised learning in cancer diagnostics. Front Oncol 2022;12:960984
24 He K, Zhang X, Ren S, Sun J Deep Residual Learning for Image Recognition. arXiv; 2015
25 Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:181004805 [cs] . 2018
26 Brown TB, Mann B, Ryder N, et al. Language Models Are Few-Shot Learners. arXiv; 2020
27 Jalali-Najafabadi F, Stadler M, Dand N, et al; BADBIR Study Group BSTOP Study Group. Application of information theoretic feature selection and machine learning methods for the development of genetic risk prediction models. Sci Rep 2021;11(01):23335
28 Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18(12):e323
29 Schünemann HJ, Cushman M, Burnett AE, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Blood Adv 2018;2(22):3198–3225

30 Key NS, Khorana AA, Kuderer NM, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: ASCO Clinical Practice Guideline Update. J Clin Oncol 2020;38(05):496–520
31 Ding R, Ding Y, Zheng D, et al. Machine learning-based screening of risk factors and prediction of deep vein thrombosis and pulmonary embolism after hip arthroplasty. Clin Appl Thromb Hemost 2023;29:10760296231186145
32 Liu Y, Song C, Tian Z, Shen W. Ten-year multicenter retrospective study utilizing machine learning algorithms to identify patients at high risk of venous thromboembolism after radical gastrectomy. Int J Gen Med 2023;16:1909–1925
33 He L, Luo L, Hou X, et al. Predicting venous thromboembolism in hospitalized trauma patients: a combination of the Caprini score and data-driven machine learning model. BMC Emerg Med 2021;21(01):60
34 Nafee T, Gibson CM, Travis R, et al. Machine learning to predict venous thrombosis in acutely ill medical patients. Res Pract Thromb Haemost 2020;4(02):230–237
35 Sheng W, Wang X, Xu W, Hao Z, Ma H, Zhang S. Development and validation of machine learning models for venous thromboembolism risk assessment at admission: a retrospective study. Front Cardiovasc Med 2023;10:1198526
36 Park JI, Kim D, Lee JA, Zheng K, Amin A. Personalized risk prediction for 30-day readmissions with venous thromboembolism using machine learning. J Nurs Scholarsh 2021;53(03):278–287
37 Qin L, Liang Z, Xie J, et al. Development and validation of machine learning models for postoperative venous thromboembolism prediction in colorectal cancer inpatients: a retrospective study. J Gastrointest Oncol 2023;14(01):220–232
38 Mantha S, Chatterjee S, Singh R, et al. Application of machine learning to the prediction of cancer-associated venous thromboembolism. Res Sq 2023:rs.3.rs-2870367
39 Rasouli Dezfouli E, Delen D, Zhao H, Davazdahemami B. A machine learning framework for assessing the risk of venous thromboembolism in patients undergoing hip or knee replacement. J Healthc Inform Res 2022;6(04):423–441
40 Nudel J, Bishara AM, de Geus SWL, et al. Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. Surg Endosc 2021;35(01):182–191
41 Katiyar P, Chase H, Lenke LG, Weidenbaum M, Sardar ZM. Using machine learning (ML) models to predict risk of venous thromboembolism (VTE) following spine surgery. Clin Spine Surg 2023;36(10):E453–E456
42 Yan YD, Yu Z, Ding LP, et al. Machine learning to dynamically predict in-hospital venous thromboembolism after inguinal hernia surgery: results from the CHAT-1 study. Clin Appl Thromb Hemost 2023;29:10760296231171082
43 Franco-Moreno A, Madroñal-Cerezo E, Muñoz-Rivas N, Torres-Macho J, Ruiz-Giardín JM, Ancos-Aracil CL. Prediction of venous thromboembolism in patients with cancer using machine learning approaches: a systematic review and meta-analysis. JCO Clin Cancer Inform 2023;7:e2300060
44 Michtalik HJ, Carolan HT, Haut ER, et al. Use of provider-level dashboards and pay-for-performance in venous thromboembolism prophylaxis. J Hosp Med 2015;10(03):172–178
45 Child S, Sheaff R, Boiko O, Bateman A, Gericke CA. Has incentive payment improved venous thrombo-embolism risk assessment and treatment of hospital in-patients? F1000 Res 2013;2:41
46 Chiasakul T, Lam BD, McNichol M, et al. Artificial intelligence in the prediction of venous thromboembolism: a systematic review and pooled analysis. Eur J Haematol 2023;111(06):951–962
47 Darzi AJ, Repp AB, Spencer FA, et al. Risk-assessment models for VTE and bleeding in hospitalized medical patients: an overview of systematic reviews. Blood Adv 2020;4(19):4929–4944
48 Chen R, Petrazzini BO, Malick WA, Rosenson RS, Do R. Prediction of venous thromboembolism in diverse populations using machine learning and structured electronic health records. Arterioscler Thromb Vasc Biol 2024;44(02):491–504
49 Ortel TL, Neumann I, Ageno W, et al. American Society of Hematology 2020 guidelines for management of venous thromboembolism: treatment of deep vein thrombosis and pulmonary embolism. Blood Adv 2020;4(19):4693–4738
50 Park HS, Ahn BJ, Jun JK. Placental transfer of clarithromycin in human pregnancies with preterm premature rupture of membranes. J Perinat Med 2012;40(06):641–646
51 Gottsäter A, Ekelund U, Melander O, Björkelund A, Ohlsson B. Cohort study of prediction of venous thromboembolism in emergency department patients with extremity symptoms. Intern Emerg Med 2024. Doi: 10.1007/s11739-024-03696-3
52 Turizo MJF, Patell R, Zwicker JI. Identifying novel biomarkers using proteomics to predict cancer-associated thrombosis. Bleeding Thromb Vasc Biol 2024;3(Suppl 1):120
53 Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? Front Surg 2022;9:862322
54 Schmuelling L, Franzeck FC, Nickel CH, et al. Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: no significant effects on report communication times and patient turnaround in the emergency department nine months after technical implementation. Eur J Radiol 2021;141:109816
55 Fisher RE, Scott JA, Palmer EL. Neural networks in ventilation-perfusion imaging. Radiology 1996;198(03):699–706
56 Jamin A, Hoffmann C, Mahe G, Bressollette L, Humeau-Heurtier A. Pulmonary embolism detection on venous thrombosis ultrasound images with bi-dimensional entropy measures: preliminary results. Med Phys 2023;50(12):7840–7851
57 Nakayama Y, Sato M, Okamoto M, et al. Deep learning-based classification of adequate sonographic images for self-diagnosing deep vein thrombosis. PLoS One 2023;18(03):e0282747
58 Soffer S, Klang E, Shimon O, et al. Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. Sci Rep 2021;11(01):15814
59 Kligerman SJ, Mitchell JW, Sechrist JW, Meeks AK, Galvin JR, White CS. Radiologist performance in the detection of pulmonary embolism: features that favor correct interpretation and risk factors for errors. J Thorac Imaging 2018;33(06):350–357
60 Das M, Mühlenbruch G, Helm A, et al. Computer-aided detection of pulmonary embolism: influence on radiologists' detection performance with respect to vessel segments. Eur Radiol 2008;18(07):1350–1355
61 Eng J, Krishnan JA, Segal JB, et al. Accuracy of CT in the diagnosis of pulmonary embolism: a systematic literature review. AJR Am J Roentgenol 2004;183(06):1819–1827
62 Abdulaal L, Maiter A, Salehi M, et al. A systematic review of artificial intelligence tools for chronic pulmonary embolism on CT pulmonary angiography. Front Radiol 2024;4:1335349
63 Rivera-Lebron B, McDaniel M, Ahrar K, et al; PERT Consortium. Diagnosis, treatment and follow up of acute pulmonary embolism: consensus practice from the PERT Consortium. Clin Appl Thromb Hemost 2019;25:1076029619853037
64 Deng RX, Zhu XL, Zhang AB, et al. Machine learning algorithm as a prognostic tool for venous thromboembolism in allogeneic transplant patients. Transplant Cell Ther 2023;29(01):57.e1–57.e10
65 Mora D, Nieto JA, Mateo J, et al; RIETE Investigators. Machine learning to predict outcomes in patients with acute pulmonary

embolism who prematurely discontinued anticoagulant therapy. Thromb Haemost 2022;122(04):570–577

66 Klok FA, Huisman MV. How I assess and manage the risk of bleeding in patients treated for venous thromboembolism. Blood 2020;135(10):724–734

67 Mora D, Mateo J, Nieto JA, et al; Registro Informatizado de Enfermedad TromboEmbólica (RIETE) Investigators. Machine learning to predict major bleeding during anticoagulation for venous thromboembolism: possibilities and limitations. Br J Haematol 2023;201(05):971–981

68 Johnstone C, Rich SE. Bleeding in cancer patients and its treatment: a review. Ann Palliat Med 2018;7(02):265–273

69 Grdinic AG, Radovanovic S, Gleditsch J, et al. Developing a machine learning model for bleeding prediction in patients with cancer-associated thrombosis receiving anticoagulation therapy. J Thromb Haemost 2024;22(04):1094–1104

70 Shahryari Fard S, Perkins TJ, Wells PS. A deep-learning approach to predict bleeding risk over time in patients on extended anticoagulation therapy. J Thromb Haemost 2024;22(07): 1997–2008

71 Choi H, Kang HJ, Ahn I, et al. Machine learning models to predict the warfarin discharge dosage using clinical information of inpatients from South Korea. Sci Rep 2023;13(01):22461

72 Ravvaz K, Weissert JA, Ruff CT, Chi CL, Tonellato PJ. Personalized anticoagulation: optimizing warfarin management using genetics and simulated clinical trials. Circ Cardiovasc Genet 2017;10 (06):e001804

73 Meid AD, Wirbka L, Groll A, Haefeli WE, Haefeli WEARMIN Study Group. Can machine learning from real-world data support drug treatment decisions? A prediction modeling case for direct oral anticoagulants. Med Decis Making 2022; 42(05):587–598

74 Zhan C, Battles J, Chiang YP, Hunt D. The validity of ICD-9-CM codes in identifying postoperative deep vein thrombosis and pulmonary embolism. Jt Comm J Qual Patient Saf 2007;33(06): 326–331

75 Fang MC, Fan D, Sung SH, et al. Validity of using inpatient and outpatient administrative codes to identify acute venous thromboembolism: the CVRN VTE study. Med Care 2017;55(12): e137–e143

76 Fanni SC, Febi M, Aghakhanyan G, Neri E. Natural language processing. In: Klontzas ME, Fanni SC, Neri E, eds. Introduction to Artificial Intelligence. Springer International Publishing; 2023:87–99

77 Lam BD, Chrysafi P, Chiasakul T, et al. Machine learning natural language processing for identifying venous thromboembolism: systematic review and meta-analysis. Blood Adv 2024;8(12): 2991–3000

78 Li A, da Costa WL Jr, Guffey D, et al. Developing and optimizing a computable phenotype for incident venous thromboembolism in a longitudinal cohort of patients with cancer. Res Pract Thromb Haemost 2022;6(04):e12733

79 Taggart M, Chapman WW, Steinberg BA, et al. Comparison of 2 natural language processing methods for identification of bleeding among critically ill patients. JAMA Netw Open 2018;1(06): e183451

80 Shung D, Tsay C, Laine L, et al. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. J Gastroenterol Hepatol 2021;36(06): 1590–1597

81 Gu K, Lee JH, Shin J, et al. Using GPT-4 for LI-RADS feature extraction and categorization with multilingual free-text reports. Liver Int 2024;44(07):1578–1587

82 Saab K, Tu T, Weng W-H, et al. Capabilities of Gemini Models in Medicine. arXiv; 2024

83 Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. arXiv:210807258 [cs] . 2021

84 Meskó B Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25: e50638

85 Wang L, Chen X, Deng X, et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. NPJ Digit Med 2024;7(01):41

86 Touvron H, Lavril T, Izacard G, et al. LlaMA: Open and Efficient Foundation Language Models. arXiv; 2023

87 Groeneveld D, Beltagy I, Walsh P, et al. OLMo: Accelerating the Science of Language Models. arXiv; 2024

88 Pecher B, Srba I, Bielikova M. Comparing Specialised Small and General Large Language Models on Text Classification: 100 Labelled Samples to Achieve Break-Even Performance. arXiv; 2024

89 Van Veen D, Van Uden C, Blankemeier L, et al. Clinical text summarization: adapting large language models can outperform human experts. Res Sq 2023:rs.3.rs-3483777

90 Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 2024;30(04):1134–1142

91 Ge J, Li M, Delk MB, Lai JC. A comparison of a large language model vs manual chart review for the extraction of data elements from the electronic health record. Gastroenterology 2024; 166(04):707–709.e3

92 Somani SS, Honarvar H, Narula S, et al. Development of a machine learning model using electrocardiogram signals to improve acute pulmonary embolism screening. Eur Heart J Digit Health 2021;3(01):56–66

93 Rovenchak A, Druchok M. Machine learning-assisted search for novel coagulants: when machine learning can be efficient even if data availability is low. J Comput Chem 2024;45(13):937–952

94 Bordukova M, Makarov N, Rodriguez-Esteban R, Schmich F, Menden MP. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. Expert Opin Drug Discov 2024;19(01):33–42

95 Ktena I, Wiles O, Albuquerque I, et al. Generative models improve fairness of medical classifiers under distribution shifts. Nat Med 2024;30(04):1166–1173

96 Mateen BA, Liley J, Denniston AK, Holmes CC, Vollmer SJ. Improving the quality of machine learning in health applications and clinical research. Nat Mach Intell 2020;2(10):554–556

97 Ng MY, Youssef A, Miner AS, et al. Perceptions of data set experts on important characteristics of health data sets ready for machine learning: a qualitative study. JAMA Netw Open 2023;6 (12):e2345892

98 Chato L, Regentova E. Survey of transfer learning approaches in the machine learning of digital health sensing data. J Pers Med 2023;13(12):1703

99 Panch T, Mattie H, Celi LA. The “inconvenient truth” about AI in healthcare. NPJ Digit Med 2019;2:77

100 Teo ZL, Jin L, Li S, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. Cell Rep Med 2024;5(02):101419

101 Arora A, Arora A. Machine learning models trained on synthetic datasets of multiple sample sizes for the use of predicting blood pressure from clinical data in a national dataset. PLoS One 2023; 18(03):e0283094

102 Coombs L, Orlando A, Wang X, et al. A machine learning framework supporting prospective clinical decisions applied to risk prediction in oncology. NPJ Digit Med 2022;5(01):117

103 Bartels R, Dudink J, Haitjema S, Oberski D, van 't Veen A. A perspective on a quality management system for AI/ML-based clinical decision support in hospital care. Front Digit Health 2022;4:942588

104 Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. Br J Radiol 2023;96(1150):20220878

105 Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28(01):31–38
106 Ferrara C, Sellitto G, Ferrucci F, Palomba F, De Lucia A. Fairness-aware machine learning engineering: How far are we? Empir Softw Eng 2024;29(01):9
107 Maccaro A, Stokes K, Statham L, et al. Clearing the fog: a scoping literature review on the ethical issues surrounding artificial intelligence-based medical devices. J Pers Med 2024;14(05):443
108 Lau BD, Haider AH, Streiff MB, et al. Eliminating health care disparities with mandatory clinical decision support: the venous thromboembolism (VTE) example. Med Care 2015;53(01):18–24
109 Lam BD, Dodge LE, Zerbey S, et al. The potential use of artificial intelligence for venous thromboembolism prophylaxis and management: clinician and healthcare informatician perspectives. Sci Rep 2024;14(01):12010
110 Bienefeld N, Boss JM, Lüthy R, et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. NPJ Digit Med 2023;6(01):94
111 Lu SC, Swisher CL, Chung C, Jaffray D, Sidey-Gibbons C. On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. Front Oncol 2023; 13:1129380
112 Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med (Lond) 2024;4(01):71
113 US Food and Drug Administration. Good machine learning practice for medical device development: guiding principles. Accessed September 26, 2024 at: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
114 White V, Nath A, Stansby G. Litigation claims relating to venous thromboembolism in the NHS. Phlebology 2015;30(06):389–396
115 Wilson E, Phair J, Carnevale M, Koleilat I. Common reasons for malpractice lawsuits involving pulmonary embolism and deep vein thrombosis. J Surg Res 2020;245:212–216
116 Hofer IS, Burns M, Kendale S, Wanderer JP. Realistically integrating machine learning into clinical practice: a road map of opportunities, challenges, and a potential future. Anesth Analg 2020;130(05):1115–1118
117 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17(01):195
118 Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. Sci Rep 2022;12(01):2398
119 TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:q902
120 Blake SR, Das N. Deploying artificial intelligence software in an NHS trust: a how-to guide for clinicians. Br J Radiol 2024;97 (1153):68–72
121 Oala L, Murchison AG, Balachandran P, et al. Machine learning for health: algorithm auditing & quality control. J Med Syst 2021;45 (12):105
122 Bates DW, Levine D, Syrowatka A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ Digit Med 2021;4(01):54
123 Cahan N, Klang E, Marom EM, et al. Multimodal fusion models for pulmonary embolism mortality prediction. Sci Rep 2023;13(01): 7544
124 Chapman BE, Lee S, Kang HP, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. J Biomed Inform 2011;44 (05):728–737
125 Banerjee I, Chen MC, Lungren MP, Rubin DL. Radiology report annotation using intelligent word embeddings: applied to multi-institutional chest CT cohort. J Biomed Inform 2018; 77:11–20
126 Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports. Radiology 2018;286(03):845–852

