

Clinical Risk Factors in Haemostasis: Evaluating VTE and Bleeding through Epidemiological and NLP-based methods

Author: Anne Bryde Alnor, MD

Department of Clinical Biochemistry, Odense University Hospital

Department of Clinical Research, University of Southern Denmark

Supervisors:

Principal Supervisor: Pernille Just Vinholt, Professor, MD, PhD

Co-supervisor: Thiusius Rajeeth Savarimuthu, Professor

Members of the Assessment Committee

Chair: Torben Bjerregaard Larsen, MD, PhD, professor

Department of Regional Health Research, Research Support Unit, Lillebaelt
University Hospital, Denmark

National reviewer: Thomas Decker Christensen, Dr.med., Clinical professor, Consultant

Department of Clinical Medicine, Department of Cardiothoracic and Vascular
Surgery, Aarhus University, Denmark

International reviewer: John Bjarne Hansen, Dr.med.

Head of Thrombosis Research Center (TREC), University Hospital of North
Norway, Tromsø, Norway

Funding:

Region of Southern Denmark

University of Southern Denmark

Department of Clinical Biochemistry, Odense University Hospital

Danish Life Science

Table of contents

Publications included in this thesis:	5
English lay summary	5
Danish lay resume	6
Thesis Rationale and Structure	8
Background	9
Haemostatic Balance and the Dual Burden of Thrombosis and Bleeding	9
Venous Thromboembolism: Risk Factors and Clinical Complexity	10
Major and Perioperative Bleeding: Challenges in Risk Identification	11
Foundations of Risk Factor Research and Clinical Application	11
Epidemiological Challenges in Bleeding and Thrombosis Research	12
Electronic Health Records and Natural Language Processing	13
Regular Expressions (Regex) and Rule-Based Approaches	14
Fundamentals of Language Models and Word Embeddings	14
Transformer-Based Models and Contextual Language Understanding	15
Combining Rule-Based Approaches with Machine Learning	16
Key objectives of this thesis:	17
Ethics	17
Statistics	17
Study presentation	20
Study I: Venous Thromboembolism Risk in Adults with Hereditary Thrombophilia: A Systematic Review and Meta-Analysis [1]	20
Study II: Extracting Perioperative Blood Loss from Unstructured Clinical Narratives	23
Study III: Natural Language Processing models in Identifying Major Bleeding Risk in Hospitalised Medical Patients [2]	26
Study IV: Early Clinical Implementation and Evaluation of an NLP-Based AI System for Thrombophilia Assessment Using Electronic Health Records	29
Discussion and Perspectives	33
Advancing Thrombotic and Bleeding Risk Prediction with NLP	33
Precision in risk stratification based on genetic variability	33
Quantification and classification of bleeding events	34
Complexity of bleeding risk in non-surgical populations	35
Integration of AI/NLP tools into clinical practice and their impact on behaviour and decision-making	35
From risk prediction to clinical implementation	36

The evolving role of NLP in Clinical Risk Assessment	37
Toward Integrated, Usable AI for Clinical Decision Support	38
Conclusion	39
References	40

Publications included in this thesis:

Study I

Venous thromboembolism risk in adults with hereditary thrombophilia: a systematic review and meta-analysis, Alnor AB, Gils C, Vinholt PJ, *Annals of Hematology* (2024) 103:4285–4294, doi.org/10.1007/s00277-024-05926-2 [1]

Study II

Extracting Perioperative Blood Loss from Unstructured Clinical Narratives, Alnor AB, Lynggaard RB, Pedersen LE, Storgaard J, Laursen MS, Vinholt PJ, under peer review.

Study III

Natural language processing for identifying major bleeding risk in hospitalised medical patients, Alnor AB, Lynggaard RB, Laursen MS, Vinholt PJ, *Computers in Biology and Medicine* (2025) May: 190: 110093, doi.org/10.1016/j.compbiomed.2025.110093 [2]

Study IV

Early Clinical Implementation and Evaluation of an NLP-Based AI System for Thrombophilia Assessment Using Electronic Health Records, Alnor AB, Adelhelm JBH, Pedersen LE, Faurø KK, Laursen MS, Lynggaard RB, Vinholt PJ, under peer review.

English lay summary

Blood clots and major bleeding are common and serious medical problems. Clots, specifically venous thromboembolisms (VTEs), can lead to life-threatening complications. On the other hand, major bleeding can be equally dangerous and unpredictable. Managing these conditions requires carefully balancing the risk of clotting against the risk of bleeding, which is often difficult because these risks change over time and vary greatly between individuals. This thesis combines traditional medical research with advanced computer methods to better understand who is at risk of VTE or bleeding.

While traditional medical research has relied on registries and structured databases, electronic health records (EHRs) contain much more detailed and accurate information on VTE and bleeding, but much of it is written in unstructured clinical notes. This means the information cannot be easily accessed or analysed using

traditional computer programs, and manually reviewing large volumes of clinical text is not realistic. To address this, an artificial intelligence (AI) method called natural language processing (NLP) enables computers to read and interpret free-text notes written by healthcare professionals, allowing relevant clinical information to be identified and extracted efficiently.

The thesis includes four studies. Study I shows that different types of hereditary thrombophilia, which are genetic mutations that increase the risk of VTE, do not carry the same level of risk. The finding challenges long-standing assumptions and highlights the need for more precise risk assessments that consider both genetic and clinical context. Study II introduces a novel method to quantify blood loss during surgery by extracting information directly from the clinical notes, helping to define what can be considered excessive bleeding for different types of operations. This is an important step toward creating better guidelines for managing bleeding during and after surgery. Study III focuses on major bleeding in hospitalised medical patients, demonstrating that major bleeding is underrecognised if relying on diagnostic codes. By using NLP, the study identified bleeding events more accurately and developed models to predict which patients were most at risk. These models showed that routine blood tests and general health status were the most useful predictors. Study IV goes beyond research and shows how NLP can be implemented in real healthcare settings. A system was developed and integrated into daily workflow to support doctors when assessing patients for thrombophilia. This tool automatically highlights relevant information from the patient's medical history, improving both accuracy and efficiency. However, Study IV also identified challenges related to clinical use, including relying too heavily on the AI system, which underlines the necessity of careful design and oversights of the tools implemented into clinical practice.

In summary, this thesis demonstrates that NLP is a valuable tool for both research and clinical practice in the context of bleeding and VTE. By extracting relevant risk factors and clinical insights from unstructured EHR text, NLP helps overcome limitations of traditional data sources and enables more accurate identification and prediction of these complications. The studies included in the thesis illustrate how NLP can support large-scale epidemiological research as well as clinical decision-making. This shows that NLP can help us better understand VTE and bleeding while also improving patient care.

Danish lay resume

Blodpropper og større blødninger er hyppige og alvorlige medicinske tilstande. Blodpropper, særligt ven blodpropper (VTE), kan føre til livstruende komplikationer. Omvendt kan større blødninger være mindst lige så farlige og uforudsigelige. Håndteringen af disse tilstande kræver en omhyggelig balance mellem

risikoen for blodprop og risikoen for blødning, hvilket ofte er vanskeligt, fordi disse risici ændrer sig over tid og varierer meget fra person til person. Denne afhandling kombinerer traditionel medicinsk forskning med metoder indenfor kunstig intelligens for bedre at forstå, hvem der er i risiko for VTE eller blødning.

Mens traditionel medicinsk forskning har baseret sig på registre og strukturerede databaser, indeholder patientjournaler langt mere detaljeret og præcis information om VTE og blødning. En stor del af disse oplysninger er dog skrevet i ustrukturerede kliniske notater. Det betyder, at informationen ikke nemt kan tilgås eller analyseres med traditionelle computerprogrammer, og manuel gennemgang af store mængder klinisk tekst er ikke realistisk. For at løse dette anvendes en metode inden for kunstig intelligens (AI) kaldet naturlig sprogbehandling (natural language processing, NLP), som gør det muligt for computere at læse og fortolke de fritekstnotater, sundhedsprofessionelle skriver. Dette gør det muligt at identificere og anvende vigtig klinisk information hurtigt og præcist.

Afhandlingen består af fire studier. Studie I viser, at ikke alle former for arvelig trombofili medfører den samme risiko for VTE. Arvelig trombofili er en fællesbetegnelse for genetiske mutationer, der øger risikoen for VTE. Studie I udfordrer tidligere antagelser og understreger behovet for mere præcise risikovurderinger, der inddrager både genetiske og kliniske forhold i en samlet vurdering. Studie II præsenterer en ny metode til kvantificering af blodtab under operationer ved at identificere oplysninger om blødning direkte i operationsbeskrivelserne. Dette hjælper med at definere, hvad der kan betragtes som en uforholdsmæssig blødning ved forskellige typer af operationer, og er derigennem et vigtigt skridt i retningen mod bedre retningslinjer for håndtering af blødning under og efter kirurgiske indgreb. Studie III fokuserer på større blødninger blandt indlagte medicinske patienter og viser, at større blødninger ofte overses, hvis man udelukkende kigger på diagnosekoder. Ved at bruge NLP kunne studiet identificere blødningsepisoder mere præcist og udvikle modeller til at forudsige, hvilke patienter der var i størst risiko for at få større blødninger. Disse modeller viste, at rutineblodprøver og patientens generelle helbredstilstand var de mest nyttige indikatorer. Studie IV tager skridtet videre og viser, hvordan NLP kan implementeres i den kliniske hverdag. Et system blev udviklet og integreret i det daglige arbejde for at støtte læger i udredningen af patienter for trombofili. Dette værktøj fremhæver relevant information i patientens journal og bidrager til øget nøjagtighed og effektivitet i udredningen. Studie IV identificerede dog også udfordringer ved klinisk anvendelse af kunstig intelligens, herunder tendensen til at stole for meget på systemet, hvilket understreger nødvendigheden af omhyggeligt design og løbende overvågning af de værktøjer, der implementeres i den kliniske hverdag.

Samlet set viser denne afhandling, at NLP er et værdifuldt værktøj i forhold til blødning og VTE både i forskningen og den kliniske hverdag. Ved at identificere relevante risikofaktorer og kliniske oplysninger fra

ustruktureret tekst i patientjournaler kan NLP omgå nogle af de begrænsninger, der er forbundet med registre og strukturerede databaser. Dette muliggør en mere nøjagtig identificering og prædiktation af større blødning og VTE. Studierne i afhandlingen illustrerer, hvordan NLP både kan anvendes i store epidemiologiske forskningsprojekter og i den kliniske hverdag. Det peger på, at NLP kan forbedre vores forståelse af VTE og større blødning og samtidig styrke kvaliteten af patientvaretagelsen.

Thesis Rationale and Structure

The identification and clinical application of risk factors for thrombosis and bleeding form the central focus of this thesis. Understanding these factors is essential to delivering safe and individualised care, particularly in patients who may require anticoagulant therapy. The thesis begins with a systematic review synthesising epidemiological evidence on venous thromboembolism (VTE) risk in individuals with hereditary thrombophilia (Study I), highlighting stratified risk across different genetic variants along with the limitations of structured data in reflecting clinical complexity.

To address these limitations, Study II investigates how bleeding-related information can be systematically extracted from unstructured clinical text using rule-based natural language processing (NLP) techniques. By quantifying perioperative blood loss at scale from surgical records, this study establishes a framework for defining pathological bleeding thresholds and supports subsequent risk factor identification.

Building upon this foundation, Study III integrates structured clinical variables with more advanced NLP approaches to detect and predict major bleeding in hospitalised medical patients. This transition marks a methodological shift from rule-based data extraction to the incorporation of transformer-based NLP models for predictive modelling of major bleeding, enabling improved risk stratification across a broader range of clinical scenarios.

Finally, Study IV translates these findings into clinical practice through the development and early evaluation of an AI-assisted decision support system. This system applies NLP to unstructured EHR text to support thrombophilia assessments, representing a significant step from retrospective modelling to prospective clinical implementation.

Collectively, the studies contribute to a more precise and data-informed understanding of risk factors for thrombosis and bleeding and their integration into clinical workflows.

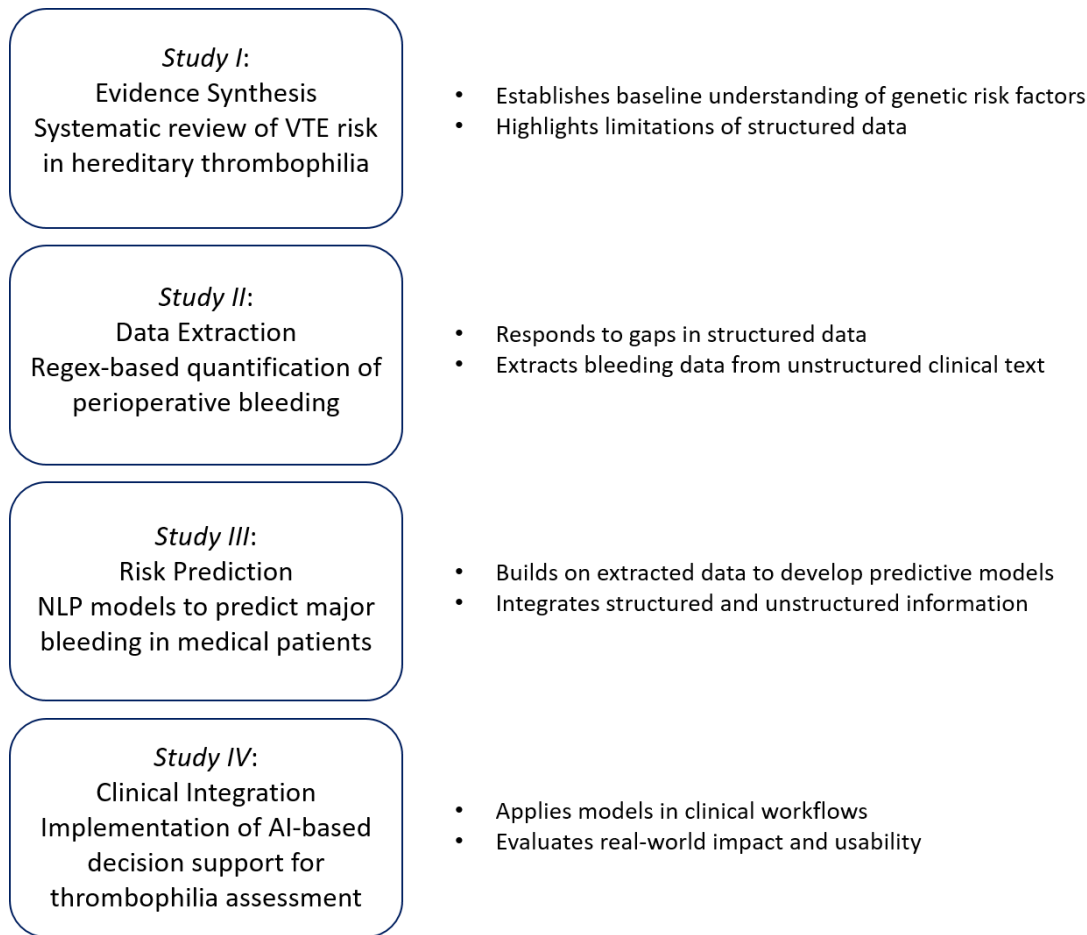


Figure 1; **Thesis Overview: A Stepwise Approach to Risk Factor Research in Haemostasis.** This figure summarises the methodological and conceptual progression across the four studies in the thesis.

Background

Haemostatic Balance and the Dual Burden of Thrombosis and Bleeding

Haemostasis is a tightly regulated physiological process that maintains blood fluidity while preventing excessive bleeding or pathological thrombosis [3]. Disruptions in haemostatic balance can result in thrombotic or haemorrhagic events, both of which are associated with substantial morbidity and mortality [4]. The medical management of haemostasis has a long history, but several challenges remain.

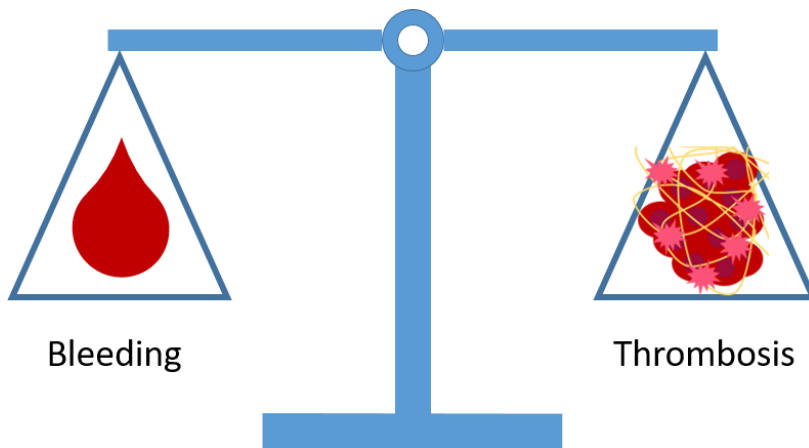


Figure 2; **Haemostatic balance.** This balance is tightly regulated under normal conditions. Disruption of this equilibrium can lead to serious clinical events such as major haemorrhage or VTE.

Venous Thromboembolism: Risk Factors and Clinical Complexity

VTE encompasses deep vein thrombosis (DVT) and pulmonary embolism (PE), and is a common and potentially life-threatening condition. The clinical presentation is often subtle and non-specific, making timely and accurate diagnosis challenging. These diagnostic difficulties highlight a broader issue: the identification of individuals at high risk for VTE remains complex, even with advances in clinical care. A key step in overcoming this challenge is recognising and understanding the diverse range of risk factors that contribute to thrombosis. The risk factors for VTE include both transient and persistent conditions each with distinct implications for clinical assessment and management. Transient risk factors, such as prolonged immobility, acute illness, trauma, surgery, and hormone therapy, often act as immediate triggers [5]. In contrast, persistent risk factors, including malignancy and hereditary thrombophilia, reflect a longer-term predisposition to thrombosis [6]. Among these, hereditary thrombophilia is particularly significant due to its genetic basis and variable clinical impact [7].

Hereditary thrombophilia refers to inherited genetic abnormalities that increase an individual's likelihood of developing thrombotic events. Several well-characterised conditions fall under this category, including factor V Leiden mutation, the prothrombin G20210A mutation, and deficiencies in natural anticoagulants such as antithrombin, protein C, and protein S. However, accurately determining the contribution of each specific genetic defect to individual VTE risk remains a challenge. This uncertainty is thought to result from a combination of factors, including variable penetrance, inconsistent expression among affected individuals, and environmental or clinical contributors [8]. Integrating genetic risk into routine decision-making requires nuanced, individualised assessments that goes beyond biochemical testing. The clinical relevance of a

positive thrombophilia test must be interpreted in the broader context of a patient's medical history and presentation to avoid misclassification or overestimation of risk.

Major and Perioperative Bleeding: Challenges in Risk Identification

Major bleeding, typically defined as bleeding that is fatal, occurs in a critical area or organ, or leads to a significant drop in haemoglobin or the need for transfusion [9], is another clinically significant yet often underrecognised complication, particularly in hospitalised medical patients. Major bleeding frequently occurs unexpectedly and is not always preceded by clearly identifiable warning signs. Risk factors for bleeding in general have been described, and include advanced age, renal or hepatic dysfunction, thrombocytopenia, active cancer, recent surgery, and the use of anticoagulants or antiplatelet agents [10, 11]. Even so, the specific risk profile for major bleeding remains poorly defined. This limits the ability to make informed decisions around anticoagulation and other haemostatic interventions.

Perioperative bleeding, though more anticipated than bleeding in medical patients, is equally complex to assess. A key difficulty lies in determining when blood loss exceeds what is expected for a given surgical procedure, and there is no universal benchmark to distinguish physiological from pathological bleeding. Establishing procedure-specific reference values is essential for developing reliable perioperative bleeding risk models. Doing so will require clinical expertise alongside analysis of large, high-quality datasets that reflect real-world variation.

Foundations of Risk Factor Research and Clinical Application

The identification and establishment of clinical risk factors are grounded in epidemiological principles that seek to distinguish causal relationships from non-causal associations. A risk factor must contribute meaningfully to the pathogenesis or progression of disease in a manner that is consistent with causality. The theoretical framework for evaluating such relationships frequently draws upon the Bradford Hill criteria, which encompass temporality, strength and consistency of association, biological plausibility, dose-response relationships, and coherence with the broader body of scientific knowledge. These criteria serve to assess whether observed associations are likely to reflect underlying causal mechanisms as opposed to confounding, bias, or random error [12].

Observational study designs are typically employed to investigate candidate risk factors when randomised controlled trials are impractical or unethical. Cohort studies facilitate the estimation of risks over time, whereas case-control studies are more efficient for exploring rare or delayed outcomes. Both designs necessitate accurate and consistent evaluation of exposures and outcomes, and appropriate control for

confounding variables. Statistical techniques such as multivariable regression, stratification, and propensity score adjustment are commonly applied to mitigate the influence of bias and isolate the independent effect of the exposure under investigation.

To strengthen the evidence of identified risk factors and to adjust for inconsistencies across individual studies, systematic reviews and meta-analyses are frequently undertaken. These methodologies offer a structured and reproducible approach to synthesising research findings, allowing for the estimation of pooled effect sizes, assessment of heterogeneity, and evaluation of potential publication bias. When conducted to a high standard, systematic reviews and meta-analyses enhance the generalisability and robustness of epidemiological evidence and often form the basis for clinical guideline development and policy formulation.

Once a risk factor has been robustly established, it may be operationalised within clinical practice through its integration into predictive models, clinical decision rules, and risk stratification frameworks. In such models, risk factors are typically weighted according to their relative contribution to the outcome of interest. These tools are intended to support, rather than supplant, clinical decision-making. By providing evidence-informed estimates of risk, they assist doctors in navigating complex scenarios.

An understanding of these conceptual and methodological foundations is essential for the critical appraisal and application of risk factor research. Nevertheless, the utility of any risk model or guideline is contingent upon the quality, completeness, and clinical relevance of the data upon which it is based. These considerations are particularly pertinent in the context of structured clinical data sources, as explored in the subsequent sections.

Epidemiological Challenges in Bleeding and Thrombosis Research

Epidemiological research has been instrumental in advancing our understanding of bleeding and thrombosis by identifying disease patterns, risk factors, and treatment outcomes. Traditional epidemiological methods have relied extensively on structured data sources, including administrative hospital records, laboratory results, and coded datasets, to quantify incidence and evaluate risk factors. While these approaches have informed clinical practice, they are limited in their capacity to reflect the complexity and heterogeneity of real-world patients.

One major limitation of structured datasets is the potential for oversimplification of complex clinical presentations, which may result in inadequate representations of patient heterogeneity. Coding systems such as the International Classification of Diseases (ICD) often lack the sensitivity and specificity required to

accurately identify cases of bleeding and VTE, leading to potential misclassification and underreporting [13, 14]. Additionally, these methods are labour-intensive and susceptible to errors. In contrast, biochemical test results, e.g. coagulation markers, renal and hepatic function tests, and haematological parameters, are typically recorded in a standardised format, making them more consistent and readily accessible for analysis. However, even these structured elements must be interpreted within a broader clinical context. These challenges highlight the need for more advanced analytical approaches that can improve data accuracy and comprehensiveness while addressing the limitations of structured epidemiological data.

Electronic Health Records and Natural Language Processing

Electronic health records (EHRs) encompass extensive clinical data, much of which exists in unstructured free-text format, including physician notes, discharge summaries, and surgical reports. While these records offer a wealth of valuable information, traditional epidemiological studies have struggled to use them effectively due to the limitations of structured data approaches. Extracting meaningful insights from unstructured clinical narratives requires more sophisticated computational techniques that can enhance the accuracy, efficiency, and depth of epidemiological research.

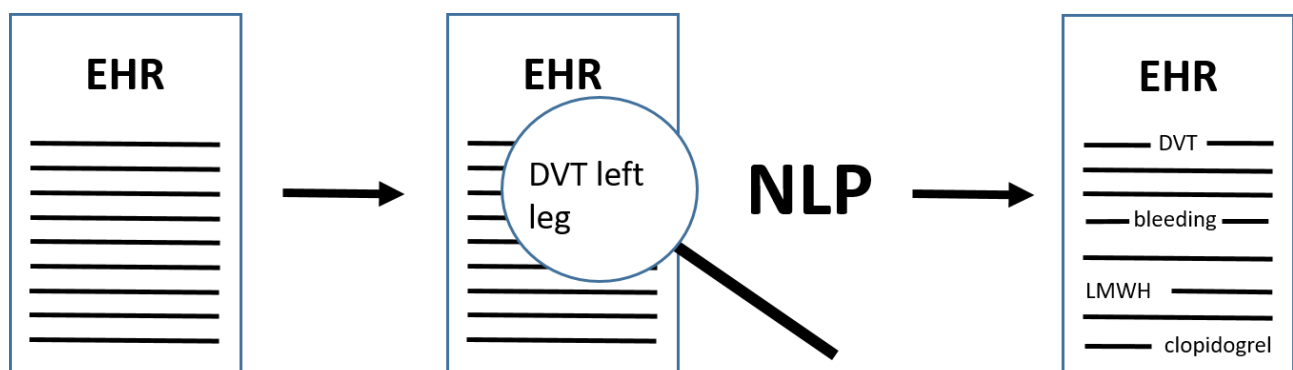


Figure 3; **NLP and the EHR**. NLP enables identification of specific clinical information from the electronic health record (EHR), such as “DVT left leg,” for structured analysis and decision support.

Technological advancements in artificial intelligence (AI) have significantly transformed medical informatics, particularly through developments in natural language processing (NLP). These improvements enable machine learning algorithms to process, interpret, and analyse vast amounts of unstructured clinical text, offering new opportunities for extracting clinical data from EHRs. The methodologies that drive these

advancements rely on both rule-based approaches, such as regular expressions (regex), and more complex machine learning-driven models, such as word embeddings and transformer-based language models.

Regular Expressions (Regex) and Rule-Based Approaches

Regular expressions (regex) is a rule-based approach to text processing that allows for pattern matching and structured data extraction from unstructured text. A regex pattern is a sequence of characters that define a search template, enabling the identification of specific substrings within larger text corpora. In the context of clinical text analysis, regex is particularly useful for extracting measured values, diagnostic terms, and pattern-based numerical markers. For example, a regex pattern can be designed to detect numerical values associated with blood loss (e.g., *Blood loss: 450 mL*), clinical parameters (e.g. *Systolic blood pressure: 134 mmHg*), or specific phrases (e.g. *Current medication Warfarin*)

The strength of regex lies in its precision. By explicitly defining patterns, it allows for the accurate extraction of known clinical terms and values, making it a powerful tool for structured data extraction. However, regex has limitations in handling linguistic variability and unseen patterns. Unlike more advanced machine learning approaches, regex lacks adaptability and generalisation capabilities. It does not infer meaning from context and is ineffective when dealing with complex sentence structures, synonyms, or misspellings. Thus, more advanced NLP models are needed to overcome the rigidity of rule-based methods.

Fundamentals of Language Models and Word Embeddings

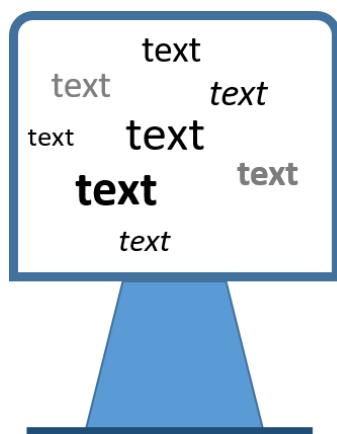
Word embeddings were developed as a more sophisticated representation of text. Instead of treating words as isolated entities, embeddings map words into a multi-dimensional space, where words with similar meanings are positioned closer together. This enables models to capture both syntactic and semantic relationships, making it possible to group related words while also distinguishing different meanings based on context. Unlike rule-based approaches, word embeddings learn these relationships automatically from large datasets, making them more adaptable to complex linguistic variations.

A commonly used approach in word embedding relies on training the models by analysing words that frequently appear in similar contexts. The core assumption is that words used in similar linguistic environments tend to share similar meanings. Word embeddings represents words as dense, continuous vectors in a high-dimensional space, positioning words with similar meanings closer together within this vector space. Through exposure to large volumes of text, the model develops a nuanced understanding of how words are used together, encoding both grammatical structures and semantic meaning.

Despite the advantages of word embeddings, they have notable limitations, particularly in handling polysemy, where a single word has multiple meanings depending on context. For instance, the word "*stroke*" can refer to a neurological condition, a swimming technique, or a gentle motion, yet traditional word embeddings assign only one fixed vector to each word. This lack of contextual differentiation poses a significant challenge in clinical NLP, where precise meaning is crucial for extracting accurate medical information from text.

Transformer-Based Models and Contextual Language Understanding

To address the challenges of word embeddings, transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) and ELECTRA (Efficiently Learning an Encoder that Classified Token Replacements Accurately) have been developed. These models go beyond simple word embeddings by introducing contextualised representations that dynamically adjust word meanings based on surrounding text. BERT uses a technique called self-attention, which helps the model understand how much each word in a sentence contributes to its overall meaning. During training, BERT learns by trying to predict words that have been deliberately masked out. In contrast, ELECTRA is trained to detect which words in a sentence have been replaced with incorrect alternatives. This makes ELECTRA more efficient, as it learns from every word in the sentence. By capturing meaning from context, the transformer models overcome key limitations of rule-based methods and word embeddings. However, as clinical language varies in structure, ambiguity, and context-dependence, no single method is sufficient on its own; combining rule-based precision with the contextual flexibility of deep learning offers a more comprehensive foundation for clinical NLP applications.



Machine learning algorithms can process, interpret, and analyse vast amounts of clinical text.

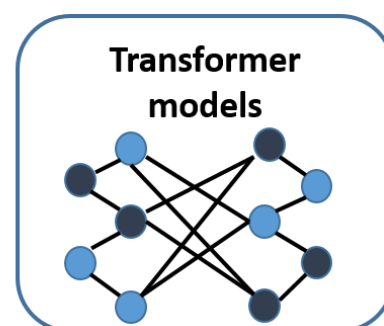
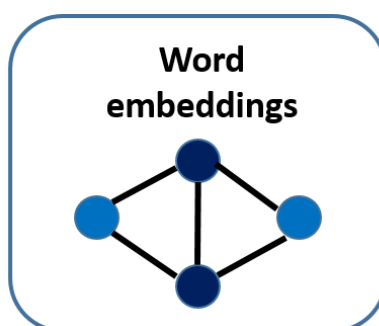
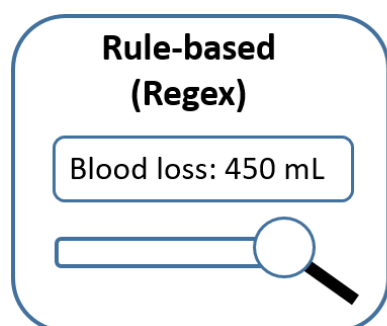


Figure 4 **Progress of NLP Techniques in Clinical Text Analysis.** The transition from rule-based regex approaches to advanced transformer-based models in extracting clinical information from unstructured HER text. While regex enables precise pattern matching, modern NLP models provide contextual understanding, supporting improved clinical decision-making and risk assessment.

Combining Rule-Based Approaches with Machine Learning

While machine learning models provide adaptability and context-awareness, they often require large amounts of labelled training data and computational resources. Conversely, rule-based methods offer precision but struggle with linguistic variability. A hybrid approach that combines the strengths of both methodologies is often the most effective in clinical NLP.

Building on this integration, this thesis aims to address the challenges of VTE and bleeding by integrating classical epidemiological methods with NLP methods to improve the understanding, risk stratification, and management of VTE and specific bleeding complications. Through systematic reviews, risk assessment tools, and language models, this work seeks to enhance predictive accuracy, inform clinical decision-making, and highlight the value of NLP-driven techniques in advancing both epidemiological and clinical research. Each study within this thesis builds on the previous, beginning with evidence synthesis, moving through

methodological development, and culminating in clinical application. Together, they form a unified framework for improving the identification and use of haemostatic risk factors in both research and practice.

Key objectives of this thesis:

1. Assess the risk of VTE in adults with hereditary thrombophilia by systematically evaluating the association between Factor V Leiden, prothrombin mutations, and deficiencies in natural anticoagulants, providing evidence to better characterise VTE risk associated with inherited thrombophilic conditions.
2. Develop and validate a regex-based model for quantifying perioperative blood loss from EHR data, to support a systematic understanding of pathological bleeding thresholds as a basis for identifying risk factors associated with excessive bleeding.
3. Apply NLP-based models to detect and characterise risk factors associated with major bleeding events in acutely hospitalised medical patients.
4. Design, implement, and conduct early clinical evaluation of an NLP-based AI system for extracting thrombosis- and bleeding-related risk factors from EHRs, with the objective of supporting thrombophilia assessment in routine clinical practice.

Ethics

Under Danish law, ethical committee approval was not required for any of the studies included in this thesis. Approval was obtained for the use of retrospective data in developing and evaluating the algorithms, with permission granted by the Region of Southern Denmark (Journal number 21/25172 and 24/43090, respectively). Permission to use retrospective data for research related to bleeding and venous thrombosis was granted by the Region of Southern Denmark under the same journal number. The Department of Clinical Biochemistry at Odense University Hospital (OUH) authorised the implementation of the NLP-based AI system in clinical practice.

Statistics

The NLP models were developed and applied using Python 3.11 (Python Software Foundation, Wilmington, USA) and the PyTorch 2.6 framework. Statistical analyses were performed using Meta-Essentials (Erasmus Research Institute of Management) [15] and STATA 18.0.

Across all studies, descriptive statistics were used to summarise patient characteristics, with normally distributed continuous variables reported as mean and standard deviation (SD) and non-normally distributed variables reported as median and interquartile range (IQR). Normality was assessed using QQ plots and the Shapiro-Wilk test. Categorical variables were compared using the Chi-square test, while non-normally distributed continuous variables were analysed using the Wilcoxon rank-sum test. Statistical significance was set at $p < 0.05$. In the following, specific statistical methods applied beyond these general approaches are detailed.

Study I [1]:

A random-effects model was employed to synthesise effect estimates across studies, with odds ratios (ORs) and corresponding 95% confidence intervals (CIs) serving as the primary effect size measure. The Mantel-Haenszel method was used for weighting individual study estimates, ensuring that each study contributed proportionally based on its sample size and variance. The results were visualised using forest plots, displaying individual study ORs alongside the overall pooled estimate.

Heterogeneity among studies was assessed using the I^2 and tau-squared (τ^2) statistics. The I^2 statistic quantified the proportion of total variability in effect estimates attributable to heterogeneity rather than chance. The τ^2 statistic, calculated via the DerSimonian and Laird method, offered additional insight into the degree of heterogeneity. A predefined threshold of $I^2 > 50\%$ and $\tau^2 > 0.50$ was used to indicate significant heterogeneity. In cases where heterogeneity was identified, subgroup analyses were performed based on study design (case-control vs. cohort studies) and risk of bias, where a low risk of bias was defined as a Newcastle-Ottawa Scale (NOS) score of ≥ 7 . Subgroup analysis based on primary and recurrent VTE were conducted regardless heterogeneity levels.

To account for expected variation in effect sizes across new studies, 95% prediction intervals (PIs) were reported, estimating the range in which the effect size of a future study from the same population would likely fall. Publication bias was assessed using both visual and statistical methods. Funnel plots were generated to examine the relationship between study effect sizes and the pooled effect size. Additionally, Egger's test was conducted to quantitatively assess funnel plot asymmetry, with a p-value < 0.05 indicating potential publication bias. Given the reduced reliability of funnel plot asymmetry assessments in small sample sizes, these analyses were not conducted for subgroups comprising fewer than ten studies.

Study III [2]

Variables that showed statistically significant differences between patients with and without major bleeding were considered for inclusion in the risk assessment models (RAMs). However, variables directly or indirectly linked to bleeding (e.g., ICU admission, number of surgical procedures, and length of hospital stay) were excluded to mitigate model bias.

Cox proportional hazards regression was used to investigate risk factors associated with major bleeding, with hazard ratios (HRs) and corresponding 95% CIs reported. The proportional hazards assumption was verified through an analysis of Schoenfeld residuals. Given the low prevalence of major bleeding in the study population, inverse probability weighting (IPW) was applied in a weighted Cox regression framework to account for potential bias. Weights were derived as the inverse of the prevalence for cases and the inverse of one minus the prevalence for non-cases. Weighted and unweighted Cox models were compared using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) to assess model fit and potential bias introduced by low event rates.

Internal validation and model performance assessment were conducted using multiple approaches. Bootstrapping with 1,000 repetitions was employed to estimate 95% confidence intervals for HRs. Harrell's C statistic was computed to evaluate model discrimination, with cross-validation performed using 10-fold cross-validation. The cross-validated area under the curve (AUC) was reported alongside its standard deviation and bootstrapped confidence intervals. Model goodness-of-fit was assessed by plotting Nelson-Aalen cumulative hazard estimates against Cox-Snell residuals.

To enhance interpretability and reduce complexity, a nested, reduced RAM was developed by retaining only variables that remained statistically significant in the multivariable analysis. The reduced and full models were compared using AIC and BIC. Predictive accuracy over time was assessed through time-dependent Brier scores, which were calculated at days 2, 4, 7, 10, 30, and 90 post-admission.

Cluster analyses were conducted to explore heterogeneity within the study population. Clusters were defined based on major bleeding type and comorbidity. The predictive performance of RAMs was evaluated within each cluster and compared to overall model performance. A minimum of ten cases per variable was required for reliable cluster evaluation. DeLong's test was employed to compare AUCs across models.

Study IV

Inter-rater reliability between NLP-assisted EHR reviews and manual reviews was evaluated using Cohen's Kappa coefficient. This measure quantifies the degree of agreement beyond chance, with values ranging from

0 (no agreement) to 1 (perfect agreement). Cohen's Kappa values between 0.40 and 0.60 were interpreted as indicating moderate agreement, while values exceeding 0.80 suggested almost perfect agreement. The agreement between NLP-assisted and manual EHR review was examined for thrombosis, bleeding, anticoagulant use, and antithrombotic use.

Study presentation

Study I: *Venous Thromboembolism Risk in Adults with Hereditary Thrombophilia: A Systematic Review and Meta-Analysis [1]*

Background and aim

Hereditary thrombophilia (HT) is a key risk factor for VTE, particularly in younger individuals with unprovoked thrombosis [16]. Although numerous studies have examined the association between HT and VTE, findings remain inconsistent [17, 18], and existing systematic reviews are either outdated [19, 20] or focused on specific subpopulations [21]. The aim of this study was to provide an updated and comprehensive evaluation of VTE risk in adults with Factor V Leiden mutation (heterozygous and homozygous), prothrombin G20210A mutation (heterozygous and homozygous), Factor V Leiden and prothrombin compound heterozygosity, and deficiencies in protein S, protein C, and antithrombin.

Methods

To achieve this, a systematic review and meta-analysis were conducted to evaluate VTE risk in adults with hereditary thrombophilia. The study adhered to the PRISMA 2020 guidelines. A comprehensive literature search was performed across PubMed, Embase (Ovid), and Web of Science to identify relevant studies. Studies included in the quantitative synthesis provided extractable data on VTE risk in adults (> 15 years) without restrictions on VTE type, location, or occurrence. Two authors independently reviewed and extracted data. The risk of bias in the included studies was assessed using the Newcastle-Ottawa Scale.

Results

The literature search identified 107 relevant publications, covering 107,130 individuals, of whom 21,560 experienced VTE. The highest VTE risk was observed in individuals with homozygous Factor V Leiden (OR 5.58, 95% CI 4.61 - 6.74), homozygous prothrombin G20210A (OR 5.16, 95% CI 3.12 - 8.52), and compound heterozygosity (OR 4.64, 95% CI 2.25 - 9.58). In contrast, the lowest risk was seen in those with heterozygous prothrombin G20210A (OR 2.21, 95% CI 1.70 - 2.87) and heterozygous Factor V Leiden (OR 2.97, 95% CI 2.41

- 3.67). Intermediate risk was noted in individuals with protein S deficiency (OR 3.01, 95% CI 2.26 - 4.02), protein C deficiency (OR 3.23, 95% CI 2.05 - 5.08), and antithrombin deficiency (OR 4.01, 95% CI 2.50 - 6.44).

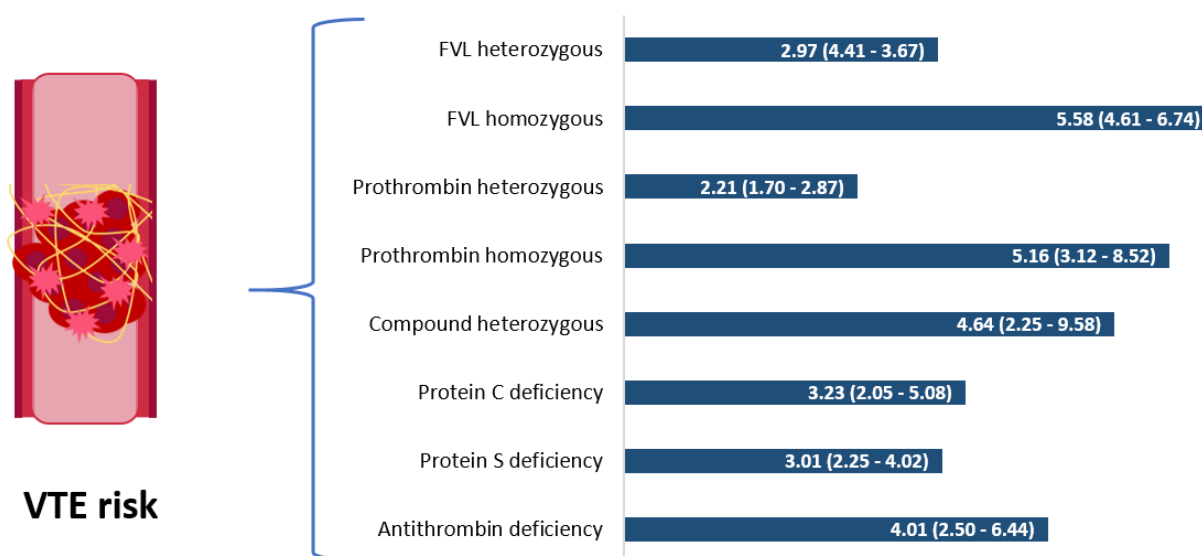


Figure 5; **VTE Risk by Type of Hereditary Thrombophilia**. Bar chart showing odds ratios and corresponding 95% confidence interval for VTE associated with different thrombophilic traits. Highest risks were seen in homozygous Factor V Leiden and prothrombin mutations, with moderate risk from anticoagulant deficiencies and lower risk from heterozygous mutations.

Conclusion

These findings confirm an elevated VTE risk in adults with hereditary thrombophilia, though the risk associated with protein C, protein S, and antithrombin deficiencies appears lower than previously reported. This discrepancy may reflect differences in the thrombotic potential of specific genetic variants. Further research is needed to refine risk stratification and improve patient management in hereditary thrombophilia.

Strengths and limitations

While this meta-analysis contributes important and up-to-date evidence on VTE risk associated with specific hereditary thrombophilias, it is essential to appraise the methodological strengths and limitations that influence the interpretation of the findings.

This study provides a contemporary and clinically relevant synthesis of VTE risk across a broad spectrum of hereditary thrombophilias. Unlike earlier reviews restricted to specific populations or subgroups, this meta-

analysis included both heterozygous and homozygous variants of Factor V Leiden and prothrombin G20210A, compound heterozygosity, and deficiencies in natural anticoagulants. This allowed for a direct comparison of genotype-specific risk estimates. Furthermore, the inclusion of all VTE manifestations, rather than focusing solely on DVT or PE, enabled a more comprehensive assessment of thrombotic risk across diverse clinical presentations.

A predefined protocol guided study selection, data extraction, and risk of bias assessment using the Newcastle-Ottawa Scale. Methodological robustness was supported by the application of a random-effects model, stratified analyses for primary and recurrent VTE, and thorough assessment of heterogeneity and publication bias using I^2 , τ^2 , funnel plots, and Egger's test. Prediction intervals were also calculated to improve the interpretability of pooled estimates in a clinical context.

However, several limitations must be acknowledged. Considerable heterogeneity was observed across studies, particularly for antithrombin deficiency and compound heterozygosity, which reduces the precision and generalisability of some estimates. Although the random-effects model accounts for between-study variability, residual heterogeneity may still influence the results. Publication bias is another concern, particularly for Factor V Leiden heterozygosity and protein C deficiency, where positive findings may be overrepresented in the literature.

Individual-level adjustment for confounders such as comorbidities, hormonal therapy, and anticoagulant use was not feasible due to reliance on aggregate data. However, these variables were systematically recorded and incorporated into the quality assessment. Inconsistent reporting of such factors across studies may still have introduced residual confounding.

To enhance internal validity, this review excluded genetically isolated populations, paediatric populations, and individuals with multiple concurrent thrombophilias. While this decision improves clarity and comparability, it may limit the applicability of the findings to certain high-risk groups. Additionally, the requirement for explicit distinction between heterozygous and homozygous states improved the specificity of the estimates but may have led to the exclusion of relevant studies.

Study I demonstrates both the value and limitations of classical epidemiological methods in assessing VTE risk. Systematic reviews and meta-analyses provide structured insights based on existing evidence but are limited by study-level heterogeneity, publication bias, and the inability to adjust for individual-level confounders. The use of aggregated data restricts consideration of clinically important factors such as

comorbidities, anticoagulant use, and hormonal therapy, potentially distorting risk estimates. These limitations highlight the need for alternative approaches that offer greater data granularity and reduced bias.

To address these limitations, Study II explores the potential of EHR-derived data, demonstrating how even simple NLP techniques can systematically extract clinically relevant information from unstructured clinical records. Using EHRs may reduce bias, enhance data completeness, and support more precise risk stratification. In this way, Study II marks a first step towards bridging the gap between traditional epidemiological methods and AI-driven clinical decision support.

Study II: Extracting Perioperative Blood Loss from Unstructured Clinical Narratives

Background and aim

Perioperative bleeding remains a significant concern in surgical and anaesthetic care, contributing to patient morbidity, increased transfusion requirements, prolonged recovery, and higher healthcare costs [22]. Current anaesthesiological guidelines emphasise the importance of estimating procedure-specific blood loss as part of preoperative bleeding risk assessment [23]. However, such data are rarely available at scale, and structured data sources, such as ICD-10 codes, often lack sensitivity, specificity, and granularity [13]. In contrast, bleeding information is often documented in free-text form within surgical and anaesthetic notes, it is not easily accessible, and thus requires a different approach than manual review or conventional epidemiological methods. The aim of this study was to develop and validate a regex-based model to extract perioperative blood loss data from unstructured clinical text.

Methods

This retrospective cross-sectional study examined patients who underwent surgery in the Region of Southern Denmark between 2013 and 2022, analysing 9.6 million surgical and anaesthesiological EHR notes from 654,855 patients. A regex-based NLP model was developed to extract both numerical (e.g., “500 mL”) and qualitative (e.g., “none”) descriptions of perioperative blood loss from unstructured EHR text. Non-numerical descriptors were assessed and, where clinically appropriate, converted to estimated volumes. To enhance sensitivity, semantically related expressions were identified using a word embedding model trained on 682 million clinical sentences, and iteratively added to the regex term set. Further refinement was achieved through manual review of 10,000 randomly sampled surgical notes with no initial blood loss extracted, pre-screened using a separate NLP model to ensure relevance. Blood loss was quantified by calculating the

median, interquartile range (IQR), and variability, defined as the range between the 1st and 99th percentiles, across surgical specialties.

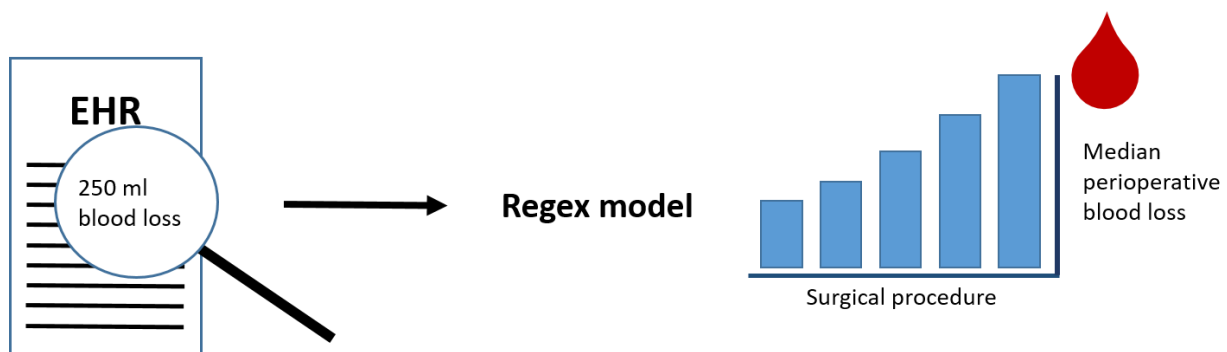


Figure 6; **Perioperative blood loss**. The regex model enabled accurate extraction of perioperative blood loss mentioned in the EHR. These data were used to calculate median blood loss across surgical procedures.

Results

The regex model demonstrated strong performance, with 99% specificity and 100% negative predictive value based on manual review of 486 surgical notes. Applied to the study cohort, the model identified quantifiable blood loss in 455,532 surgical notes from 196,342 patients. The overall median blood loss was 103 mL (IQR 0–350 mL), with obstetric surgeries showing the highest median (300 mL, IQR 200–484 mL). Four surgical groups had a median of 0 mL. Considerable variability was observed across specialties, with cardiothoracic, vascular, and gastrointestinal procedures each showing a >4,000 mL range between the 1st and 99th

percentiles.

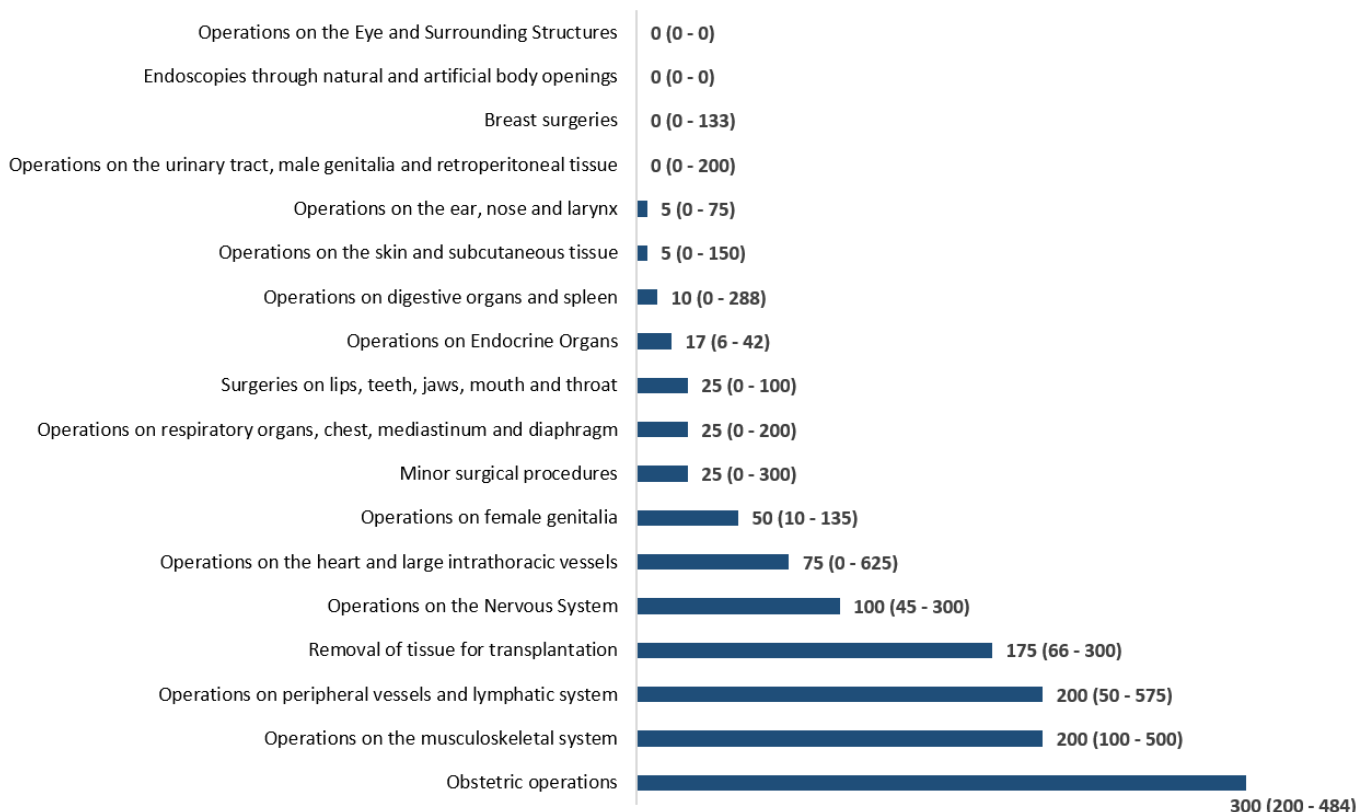


Figure 7; **Median perioperative blood loss by surgical speciality.** The figure shows the median and interquartile range of perioperative blood loss according to surgical speciality.

Conclusion

The regex-based model effectively extracted perioperative blood loss from unstructured EHR text, providing insights into variability across surgical procedures and aiding evidence-based perioperative management.

Strengths and limitations

This study demonstrates the feasibility of extracting perioperative blood loss data from a large-scale, real-world dataset using a regex-based approach. By integrating both quantitative values and qualitative descriptions, the method enhances the completeness and precision of bleeding assessment. The model showed strong validation performance, supported by a structured development process involving manual review and iterative refinement. A key strength of the approach lies in its ability to structure previously unstandardized information at scale. This provides a necessary empirical foundation for defining excessive

blood loss and supports future research into associated risk factors and the development of more context-specific perioperative risk models. The identification of substantial variability in procedures highlights the need for procedure-specific risk assessments.

However, the study has limitations. The findings are based on a single healthcare region, and differences in documentation practices across hospitals and countries may affect generalisability, necessitating further validation in other settings. Moreover, the study does not assess clinical outcomes such as transfusion requirements or haemodynamic instability. Integrating this model with structured clinical data could enhance its utility for perioperative risk assessment and patient management.

Reporting bias is another concern, as quantifiable blood loss was documented in only one-third of patients, potentially leading to an overrepresentation of significant bleeding cases. Furthermore, the rule-based approach depends on predefined patterns, which may not capture all variations in how blood loss is documented. While manual review helped mitigate this issue, some bleeding events may still have been missed.

While Study II demonstrates the feasibility of using rule-based NLP methods to extract perioperative blood loss from unstructured EHR data, it also highlights the limitations of fixed-pattern approaches in adapting to language variation and incomplete documentation. Building on this, Study III transitions from descriptive extraction to predictive modelling, integrating clinical text with biochemical markers and comorbidities to develop risk assessment models for bleeding. This progression marks a shift from structured data capture to dynamic risk estimation, illustrating the broader potential of NLP for real-world clinical decision support.

Study III: Natural Language Processing models in Identifying Major Bleeding Risk in Hospitalised Medical Patients [2]

Background and aim

Major bleeding is a serious yet often overlooked condition in medical patients, largely due to the complexity of its assessment. Determining whether a bleeding event qualifies as major requires integrating multiple clinical data points, making large-scale retrospective analysis highly challenging [9]. This study addresses this limitation by applying an advanced NLP model, developed using word embeddings with a transformer-based encoder, to identify bleeding events. Simpler rule-based algorithms are used alongside it to extract potential

risk factors. The findings from these models are then integrated into risk assessment models (RAMs) to enhance the prediction of major bleeding in hospitalised medical patients.

Methods

This retrospective cross-sectional study analysed EHRs of adult patients admitted through the Emergency Department at Odense University Hospital between January 2017 and December 2022. Bleeding events during admission were identified using an NLP model developed to identify all types of bleeding within unstructured Danish EHR text. Definition of major bleeding was based on current International Society of Thrombosis and Haemostasis guidelines [9] after manual review of bleeding events. Risk factors such as alcohol consumption, smoking status, blood pressure, BMI, anticoagulant or antithrombotics use were identified through a combination of regex search patterns and a transformer encoder model. Demographics, comorbidities, and admission biochemical values were also assessed. A comprehensive and a simplified RAM were developed and evaluated. The comprehensive model incorporated all relevant information, including anamnestic data, while the simplified model focused on biochemical parameters, improving its practicality for real-world applications.

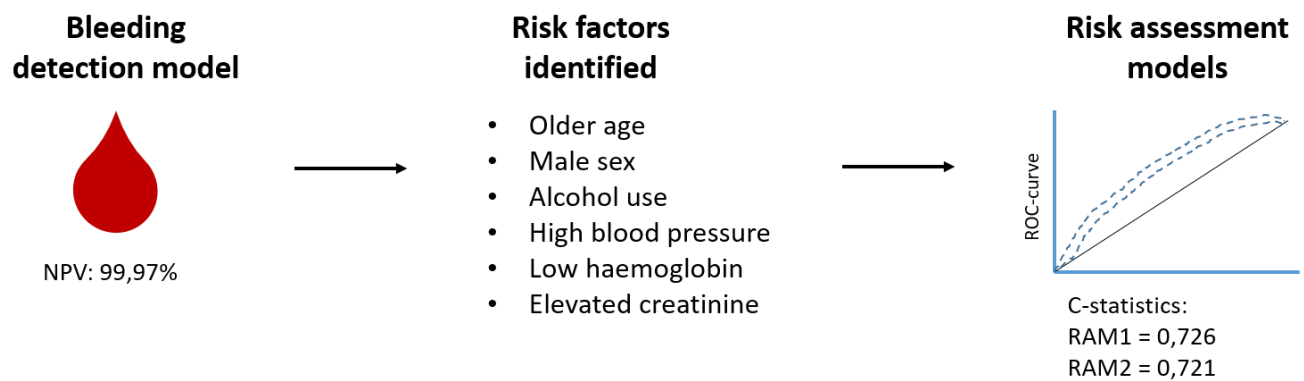


Figure 8; **NLP in major bleeding assessment.** NLP was used to identify bleeding episodes in the EHR. Several anamnestic and biochemical risk factors for major bleeding were identified, and the resulting risk assessment models (RAMs) demonstrated strong predictive performance.

Results

The bleeding detection model demonstrated a negative predictive value of 99.97%, ensuring no major bleeding cases were missed. Smoking and alcohol classification achieved a positive predictive value of 98%

at both the sentence and patient level, while blood pressure and BMI extraction showed 100% accuracy. Of 46,439 eligible patients, 1,246 (2.7%) experienced major bleeding. Identified risk factors included older age, male sex, alcohol consumption, higher systolic blood pressure, lower haemoglobin, and elevated creatinine. RAM 1, incorporating biochemical data and comorbidities, showed strong predictive performance (Harrell's C-statistic = 0.726). RAM 2, a simplified model excluding comorbidities, achieved similar accuracy (C-statistic = 0.721), suggesting its applicability in real-world settings. Both models remained consistent across validation.

Conclusion

This study identifies the incidence and risk factors of major bleeding in medical patients, highlighting the predictive value of routinely measured biochemical markers. It also demonstrates the effectiveness of NLP models in detecting clinical information within EHR text.

Strengths and limitations

This study has several methodological strengths. The use of NLP methods to extract bleeding-related data from EHRs enhances sensitivity and specificity in detecting bleeding events. This approach reduces limitations associated with traditional data collection and provides a more detailed assessment of bleeding risk factors.

The large dataset, comprising 46,439 medical patients, ensures robust statistical power. The integration of Denmark's national health registers and laboratory information systems enables comprehensive data collection on patient demographics, comorbidities, biochemical parameters, and clinical outcomes. To maintain focus on non-surgical medical patients, those undergoing surgery due to bleeding were excluded.

The study objectively identifies risk factors for major bleeding by incorporating routinely measured biochemical parameters, such as haemoglobin, CRP, albumin, and creatinine, rather than relying solely on clinical assessments or patient history. These variables provide actionable targets for risk assessment in clinical practice.

The predictive models were developed and validated using Cox proportional hazards regression, cross-validation, bootstrapping, and Brier scores. Their performance was comparable to or exceeded that of existing bleeding risk models. Time-dependent predictive accuracy was assessed at multiple time points (2, 4, 7, 10, 30, and 90 days) to ensure relevance throughout a patient's hospital stay. Cluster analyses further demonstrated model robustness across different types of major bleeding and comorbidities.

Despite these strengths, the study has limitations. Its retrospective and observational design prevents causal inference, and while statistical adjustments mitigate biases, prospective validation is needed. The exclusion of patients with severe cardiac, haematological, and oncological diseases limits generalisability to these high-risk groups. Self-reported data on smoking and alcohol consumption may introduce information bias, and incomplete records further complicate risk estimation.

Selection bias is another consideration, as the study includes only patients admitted via the emergency department, making findings less applicable to outpatient settings or specialised inpatient units. Additionally, while the NLP model effectively extracts bleeding-related data, its accuracy depends on the quality of clinical documentation in EHRs.

Challenges in predictive modelling were also noted. While RAM 1 showed slightly stronger predictive performance, RAM 2 offered a more streamlined alternative with similar efficacy and lower complexity, highlighting a trade-off between model performance and clinical utility. Attempts to use inverse probability weighting (IPW) to adjust for biases resulted in violations of the proportional hazards assumption, suggesting that alternative methods may be required for further model refinement.

While Study III demonstrates the utility of advanced NLP models in detecting major bleeding and identifying associated risk factors, its focus remained on retrospective analysis and risk prediction. Building on this, Study IV moves from retrospective stratification to clinical implementation, applying NLP models prospectively to extract risk factors essential for thrombophilia assessment. This marks a critical step towards real-time clinical decision support, where AI-assisted systems are integrated directly into diagnostic workflows to support clinicians, improve efficiency, and enhance patient care.

Study IV evaluates the practical use of these tools in a clinical setting, assessing their impact on workflow, diagnostic accuracy, and doctor acceptance. Unlike the retrospective design of the preceding studies, Study IV adopts a prospective clinical approach, demonstrating how NLP-based systems can transition from theoretical development to real-world application in support of personalised thrombosis risk evaluation.

Study IV: Early Clinical Implementation and Evaluation of an NLP-Based AI System for Thrombophilia Assessment Using Electronic Health Records

Background and aim

Targeted testing for thrombophilia is recommended for select patients to guide treatment and risk assessment, but interpreting thrombophilia requires integrating biochemical and anamnestic data [24]. The unstructured nature of anamnestic information in EHRs makes manual review time-consuming and error-prone. This study aimed to develop, implement, and evaluate a NLP-based AI system designed to identify and extract clinically relevant anamnestic information to support thrombophilia assessments in routine clinical care.

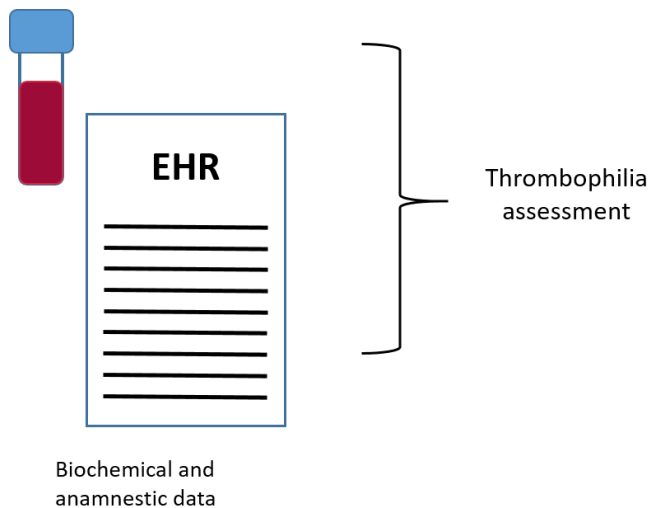


Figure 9; **Elements of a thrombophilia assessment.** Accurate evaluation of thrombophilia and the associated risk of VTE requires careful integration of both anamnestic and biochemical information to guide appropriate anticoagulant treatment.

Methods

The AI system was developed using NLP models trained on 299,718 Danish EHRs from OUH spanning 2015 - 2022, encompassing 1.4 billion tokens [25]. These models were fine-tuned on clinical narratives to identify sentences related to thrombosis, bleeding events, anticoagulant use, and antithrombotic therapy. Pre-implementation validation on an independent dataset demonstrated strong performance (NPV > 99%, PPV ≥ 95%). The AI system was integrated into the thrombophilia assessment workflow, allowing doctors to use its structured output either as a supplement to or a replacement for manual EHR review. Its performance was evaluated by comparing AI-assisted reviews with traditional manual reviews in a retrospective cohort of 50 patients assessed for thrombophilia. Additionally, usability was examined through an eye-tracking study assessing review time, accuracy, and information retrieval patterns. Semi-structured interviews with doctors were conducted to evaluate feasibility, workflow integration, and potential adoption barriers.

Results

The AI system was successfully integrated into routine care. In the retrospective cohort, AI-assisted reviews demonstrated high agreement with original documentation, identifying all previously recorded findings and uncovering additional relevant information in 68% of cases. Inter-rater reliability for thrombosis findings was excellent (Cohen's $\kappa = 0.92$), with strong agreement also observed for anticoagulant ($\kappa = 0.75$) and antithrombotic ($\kappa = 0.82$) domains. Bleeding-related findings showed moderate agreement ($\kappa = 0.58$), although AI-assisted reviews captured more bleeding events than were originally documented.

Eye-tracking analysis further demonstrated that AI assistance improved efficiency in navigating complex EHRs. Review time was reduced by a median of 3.33 minutes (up to 33.3% faster), and relevant information was located 35.2% more quickly. The eye-tracking data revealed two distinct usage patterns: active filtration, where doctors relied on AI-generated highlights to navigate the record, and manual review through traditional scrolling. Active filtration was associated with faster identification of relevant content.

Doctors reported high trust in the AI system's accuracy and data security. They highlighted clear advantages, including improved information retrieval and a better overview in complex cases. Reported barriers to broader adoption included limited EHR integration, suboptimal alignment with clinical workflows, and the need for stronger institutional support. No system errors, malfunctions, or patient safety concerns were reported during the implementation period.

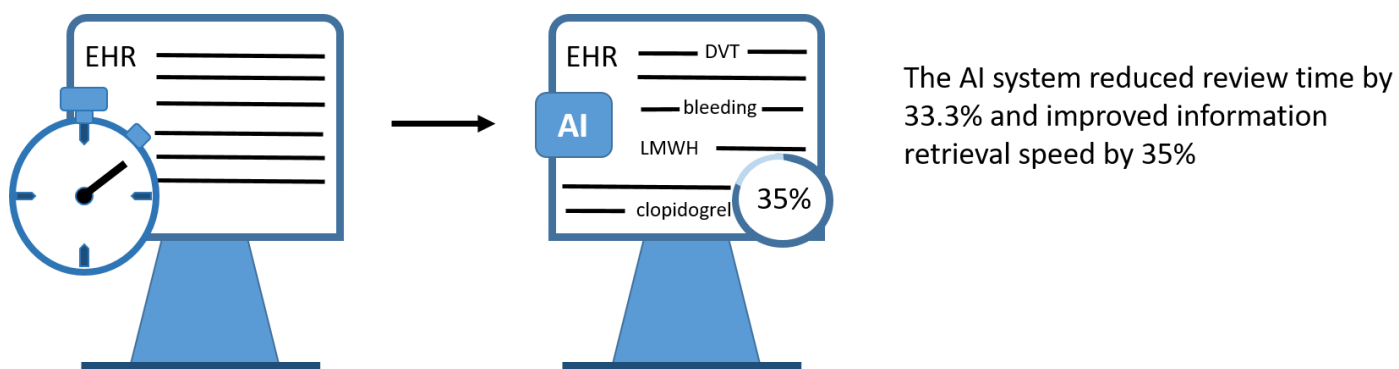


Figure 9; **AI-Assisted Highlighting of Relevant Clinical Information in EHRs.** The AI system processes unstructured EHR text and highlights key clinical concepts including thrombosis, bleeding events, and medications (e.g. LMWH, clopidogrel) to support thrombophilia assessment and clinical decision-making.

Conclusion

This study demonstrates the clinical utility of a real-world AI system that uses NLP to identify relevant patient history from unstructured EHR text. The system enhanced completeness and accuracy of thrombophilia assessments, improved review efficiency, and was well-received by doctors. While not a replacement for clinical judgment, it serves as an effective assistive tool in supporting decision-making. Future efforts should focus on integrating the system directly within the EHR, mitigating automation bias through structured training, and assessing long-term impacts on clinical outcomes.

Strengths and limitations

This study presents a comprehensive development, validation, and implementation of an AI-assisted decision-support system for thrombophilia assessment, covering the full translational pipeline from model training to clinical deployment. By validating the thrombosis model across multiple EHR platforms and applying it in routine care, the study demonstrates robustness across technical environments and relevance to real-world clinical practice. The modular NLP architecture, incorporating domain-specific models for thrombosis, bleeding, anticoagulants, and antithrombotics, supports clinically meaningful information extraction from unstructured text.

A key strength is the triangulated evaluation approach, combining technical performance metrics, simulation-based eye-tracking of information-seeking behaviour, and qualitative analysis of clinician adoption based on a recognised implementation framework. This multi-dimensional assessment captures both system accuracy and behavioural impact, offering insights into cognitive workflow integration and human-AI interaction. The use of sentence-level classification with interpretable outputs enhances transparency, while validation on both selected and unselected populations supports applicability across diverse clinical contexts. Overall, the study contributes methodologically rigorous and practically grounded evidence for the integration of AI in complex decision-making processes involving unstructured EHR data.

However, several limitations should be noted. Conducted at a single institution, the findings may not be directly applicable to healthcare systems with different EHR infrastructures or documentation practices. While the AI system was validated on a large retrospective dataset, the prospective implementation was limited, which may limit broader applicability. Further research across multiple institutions and diverse patient populations is needed to confirm its external validity.

The study employed eye-tracking to assess doctor interaction with the AI system, though this method has inherent limitations in capturing cognitive processes. While results indicate that AI-assisted reviews improved focus on relevant clinical data, the direct impact on decision-making and patient outcomes remains unclear.

Future studies incorporating real-time decision-making assessments and follow-up data would provide stronger evidence of clinical benefit.

Automation bias was another key consideration, as eye-tracking data indicated that doctors sometimes overlooked unmarked findings, highlighting the need for structured training to ensure critical evaluation of AI-generated outputs. The study also observed variability in usage patterns, suggesting that workflow integration and onboarding strategies require further refinement.

Additionally, as this study focused on short-term implementation, the long-term effects on diagnostic accuracy, clinical decision-making, and patient outcomes remain unknown. Further research should explore whether AI integration reduces diagnostic errors, improves treatment decisions, or enhances patient care over time. Considerations around cost-effectiveness and resource allocation are also necessary for assessing feasibility in routine practice.

Overall, this study demonstrates the feasibility of an AI-powered NLP system for thrombophilia assessment, enabling efficient data extraction from unstructured EHRs with high accuracy and doctor acceptance. While challenges such as workflow adaptation and automation bias remain, real-world implementation confirms the system's potential to streamline decision-making without compromising safety.

Discussion and Perspectives

Advancing Thrombotic and Bleeding Risk Prediction with NLP

Understanding VTE and bleeding risk has traditionally relied on structured epidemiological methods, often constrained by limited granularity and retrospective biases. The findings in this thesis demonstrate that modern NLP approaches can enhance our ability to detect, quantify, and predict thrombotic and haemorrhagic complications. By bridging classical statistical methodologies with AI-driven analysis of EHRs, this research highlights a crucial transition from retrospective association studies towards a dynamic clinical decision support.

Precision in risk stratification based on genetic variability

Study I provides updated, stratified VTE risk estimates for individuals with hereditary thrombophilia, emphasising considerable variability and significant heterogeneity within traditionally defined risk groups, particularly regarding natural anticoagulant deficiencies (Protein C, Protein S, and Antithrombin). This

challenges the widely held assumption of uniformly high thrombotic risk associated with these deficiencies [26, 27], suggesting that individual genetic variants within each deficiency are associated with distinctly different risk levels. Study I emphasises the importance of evaluating VTE risk through an integrated clinical approach that considers both patient history and the specific characteristics of the identified thrombophilia. Risk is not solely determined by the presence of a biochemical abnormality but varies substantially within thrombophilia types, depending on the underlying genetic variant and its phenotypic expression. Natural anticoagulant deficiencies such as antithrombin, protein C, or protein S deficiency represent heterogeneous conditions, and a single risk estimate cannot be applied uniformly across all genotypes. Treatment decisions should therefore be based on a nuanced interpretation that integrates molecular subtype with clinical context. This approach supports more precise and individualised management, and improves the relevance and quality of genetic counselling for patients and their families [28].

Quantification and classification of bleeding events

Study II addresses a key clinical need for accurate, large-scale quantification of perioperative blood loss, which is essential for bleeding risk assessment and surgical planning. By quantifying both median blood loss and interpatient variability, the findings support a more detailed understanding of bleeding patterns, which may inform perioperative management strategies, transfusion planning, and clinical guideline development. The marked variability in bleeding both within and across surgical specialties underscores the need for individualised perioperative planning. However, this variability also highlights a considerable limitation in current clinical practice and research, namely the lack of a standardised, clinically meaningful definition of excessive, or major, surgical bleeding [29]. Existing definitions are inconsistent and often fail to consider the expected blood loss for a given procedure, hindering the comparability of studies and the development of universal guidelines. This limitation is also reflected in the recent joint consensus statement [30], which highlights the absence of a universal definition of major bleeding in cardiac surgery and criticises current quality metrics, such as re-exploration rates, as inadequate. The statement advocates for more comprehensive, standardised bleeding definitions to support quality benchmarking and improve clinical outcomes across institutions [31]. Establishing procedure-specific thresholds is an essential first step towards identifying risk factors associated with excessive perioperative bleeding. Without a reliable definition of what constitutes excessive bleeding, it is not possible to meaningfully determine which clinical characteristics, interventions, or patient factors contribute to elevated risk.

Complexity of bleeding risk in non-surgical populations

Study III focuses on identification and stratification of major bleeding risk among acutely hospitalised medical patients, who are underrepresented in bleeding risk models. Traditional risk models rely on predefined variables and structured data [10, 32]. However, clinical decision-making often involves unstructured, narrative patient data that cannot be easily captured using standardised coding systems. The use of AI-enhanced models allows for the extraction of clinically relevant information from free-text medical records, offering a more nuanced and patient-specific approach to risk stratification [33]. Similarly, it has been shown, that AI-driven haemostasis risk stratification models outperformed traditional scoring systems by integrating real-time haemodynamic data with clinical notes [34].

Study III found that only 39% of major bleeding events were identified through ICD-10 codes, aligning with emerging evidence that highlights the limitations of administrative coding in accurately capturing bleeding events [13, 35]. These findings further support a growing shift away from sole reliance on ICD-10 codes toward more comprehensive, data-driven approaches for clinical event identification [31]. Study III moreover challenges prevailing assumptions that anticoagulant use is the primary predictor of bleeding [36]. Instead, it found that biochemical markers and broader patient-level factors, such as comorbidity burden and alcohol use, were stronger predictors. A recent study found that concomitant use of anticoagulants and antithrombotics did not significantly increase the risk of major bleeding, supporting this shift in emphasis [37]. This perspective is supported by a systematic review that showed that patients not receiving anticoagulant therapy for secondary VTE prophylaxis still experienced major bleeding events [38]. Together, these findings suggest that bleeding risk in general medical inpatients may be more closely tied to overall physiological frailty than to anticoagulation status alone. NLP models provide an effective means of identifying clinical indicators of frailty that would otherwise require time-consuming manual review of the EHR or risk being omitted when relying solely on structured data.

Integration of AI/NLP tools into clinical practice and their impact on behaviour and decision-making

Study IV focuses on the clinical application of an NLP-based AI system designed to support thrombophilia assessments by highlighting relevant information in unstructured EHR text. Using a combination of eye-tracking analyses and qualitative interviews, the study explores clinician use and perception of the AI system in routine practice. The eye-tracking component highlights behavioural patterns in information retrieval, offering a deeper understanding of how AI assistance influences clinical review strategies. Active filtration, i.e. where clinicians relied primarily on AI-generated highlights to navigate the record, was associated with the greatest reductions in review time. This suggests that users increasingly adapted their behaviour to the

structure and suggestions provided by the AI, streamlining their engagement with the record and focusing attention on highlighted elements. While this approach may improve efficiency, particularly in time-pressured clinical environments, it also introduces potential risks. The findings suggest a degree of automation bias, reflecting a tendency to focus on AI-highlighted content at the expense of non-highlighted information. This aligns with concerns raised in previous studies, which have shown that although digital tools can enhance clinical efficiency, they may also contribute to information neglect when clinicians place uncritical trust in algorithmic outputs [39-41]. Study IV shows that the value of AI support lies not only in its technical accuracy, but in how it is used within clinical decision-making. Addressing this requires not just thoughtful implementation and training, but also attention to broader usability issues. In line with previous research showing that prolonged and inefficient EHR interaction can contribute to physician fatigue and reduced performance [42], this study suggests that AI tools, when used appropriately, may help alleviate cognitive burden by improving information accessibility and streamlining review processes. Crucially, these issues of cognitive strain and automation bias are often rooted in poor EHR navigation design. As previously noted, ineffective navigation remains a key threat to both usability and patient safety [43]. By automating the identification and surfacing of clinically relevant history, the AI system evaluated in this study provides a targeted response to these longstanding challenges, thereby offering a practical example of how intelligent navigation support can improve both the safety and efficiency of clinical documentation review.

These implementation-related findings also reflect broader themes in the literature on integrating advanced functionalities within EHR systems. For instance, the importance of using structured implementation science frameworks to guide the integration of EHR-based tools into learning health systems has been emphasised [44]. Findings from Study IV converge with previous studies on several key factors that influence adoption: the perceived clinical utility of the tool, its compatibility with existing workflows, ease of use, and the visibility of its benefits in day-to-day practice [45]. Together, these studies underscore the critical need to align technological innovation with clinician experience and organisational context to ensure sustainable, safe, and meaningful use of digital tools in healthcare.

From risk prediction to clinical implementation

This thesis provides an example of how NLP methods are progressing from conceptual development to practical use in clinical settings. While Study I employed classical meta-analytic methods to assess VTE risk in hereditary thrombophilia, Studies II and III moved beyond structured data analysis by applying NLP techniques to extract bleeding-related information directly from EHRs. This methodological shift reflects an increasing emphasis on using unstructured clinical data to support more nuanced risk assessment.

The findings align with broader efforts to apply NLP for real-time patient risk monitoring, including assessing patient risk of cardiac disease [46] and opioid misuse [47] from EHR data. Within the VTE domain, NLP has largely been used to retrospectively detect thrombotic events, often relying on radiology reports [48] or validating performance against structured coding systems [49]. While these studies have demonstrated high accuracy, and some systems have been evaluated for near real-time use [50-52], they typically remain focused on event detection. More recently, NLP components have been incorporated into clinical decision support systems to assist with NLP identification and classification [53], though their integration into broader clinical reasoning remains limited.

Study IV extends the use of NLP into a more complex and clinically nuanced domain. The system developed for thrombophilia assessment was designed to identify relevant anamnestic risk factors from free-text EHR entries. In contrast to previous approaches, this study focused on actual clinical implementation, addressing issues such as usability, clinician trust, automation bias, and compatibility with existing workflows. Most importantly, the study repositioned NLP from a tool for retrospective event detection to a system that supports longitudinal clinical decision-making.

This development reflects broader trends in the field, where NLP is increasingly combined with predictive models to enable clinical risk assessment and decision support [54-56]. Recent studies have shown how NLP can be used to identify patients at risk of gastrointestinal bleeding [57], extract features for stroke classification [58], and detect conditions such as heart failure [59], sepsis [60], and medication-related adverse events [61]. These approaches range from keyword matching and regex model to deep learning architectures capable of capturing complex clinical context. Collectively, these studies illustrate the expanding role of NLP in supporting both standalone applications and integrated clinical workflows across multiple specialities. A key insight from Study IV is the emphasis on integrating NLP systems within clinician-facing interfaces that prioritise usability and transparency. Features such as clickable highlights, structured outputs linked to sources text, and intuitive interface design are essential not only for building trust, but also for supporting clinical verification and ensuring safe, effective implementation in routine care.

The evolving role of NLP in Clinical Risk Assessment

Risk stratification is central to clinical decision-making in haemostasis, where accurate assessment of bleeding and thrombotic risk guides treatment selection, monitoring strategies, and preventive interventions. Traditional risk models, such as the Wells score for venous thromboembolism [62] or HAS-BLED for bleeding [63], are based on a limited number of predefined variables and are typically tailored to specific patient populations. While these tools are widely used and validated, they provide static assessments

that may not reflect the complexity or temporal variability of patient care. Additionally, there are persistent concerns regarding their generalisability and potential biases, especially when applied to populations differing from those in which the models were originally developed.

As healthcare data becomes increasingly complex and digitised, there is growing interest in developing dynamic, AI-driven risk models that integrate a broader range of clinical variables. These include structured data such as vital signs, laboratory results, and comorbidities, as well as unstructured data from clinical notes. The future of risk modelling in haemostasis may lie in adaptive systems that continuously update individual risk profiles. Embedding such models within real-time decision support tools could enable dynamic risk assessments that reflect ongoing changes in a patient's condition. By incorporating haemodynamic trends, laboratory data, and clinical narratives, such systems would have the potential to identify elevated bleeding or thrombotic risk before adverse events occur, supporting more proactive and individualised care. Nevertheless, ensuring fairness, transparency, and clinical relevance remains a significant challenge in the development and implementation of such technologies.

Toward Integrated, Usable AI for Clinical Decision Support

Despite the promising results demonstrated in this thesis, several challenges remain before such systems can be widely adopted in routine clinical practice.

First, explainability remains a major barrier to trust and adoption. While rule-based NLP models, such as regex approaches used in Study II, provide transparent extraction processes, deep learning models (e.g., transformers used in Study III) function as “black boxes,” making it difficult for doctors to understand how predictions are generated. There is a growing consensus in AI research that explainable AI should be prioritised to enhance doctor confidence in automated systems [64]. Future work should focus on integrating interpretable machine learning methods that provide clear rationales for risk predictions.

Second, workflow integration is a critical factor for AI adoption. Study IV demonstrated that AI-assisted thrombophilia assessments improved efficiency, yet doctor feedback indicated that adapting to new digital tools remains a challenge. While several studies have shown an overall positive attitude towards the implementation of AI in decision support context [65, 66] scepticism remains, particularly when comparing and AI-assisted output with a human evaluation [67, 68]. Addressing this issue requires user-centred design principles, ensuring that AI tools are seamlessly integrated into existing clinical workflows without disrupting standard practice.

Finally, regulatory considerations must be addressed before AI models can be widely deployed in patient care. Unlike traditional risk assessment tools, AI-driven models require continuous monitoring, updating, and validation to maintain accuracy and safety. The European Medicines Agency has recently introduced guidelines on AI in healthcare, emphasising the importance of rigorous validation and real-world performance monitoring before widespread implementation [69]. Future research should prioritise regulatory-compliant AI development, ensuring models meet ethical and legal standards while maintaining clinical utility.

Therefore, while real-time AI-driven decision support tools offer clear potential to improve haemostatic risk stratification and clinical outcomes, their successful adoption will require advances in interpretability, thoughtful integration into clinical practice, and robust regulatory compliance. By addressing these challenges, future systems may move beyond theoretical models to become essential, trusted tools in personalised patient care.

Conclusion

This thesis explored risk factors associated with thrombosis and bleeding across a series of studies that reflect a methodological progression from traditional epidemiological approaches to the application of NLP models. Drawing on both classical methods, such as systematic evidence synthesis, and more recent computational techniques, including rule-based text extraction and deep learning, the work examined how structured and unstructured clinical data can be used to enhance risk assessment and support clinical decision-making.

The first study systematically evaluated the risk of VTE in adults with hereditary thrombophilia by assessing associations between Factor V Leiden, prothrombin G20210A, and deficiencies in natural anticoagulants (protein C, protein S, and antithrombin). The findings highlighted notable variation in VTE risk across different genetic subtypes, challenging the notion of uniformly high risk. This contributes to a more differentiated understanding of thrombophilia and supports improved risk stratification, clinical management, and genetic counselling.

The second study involved the development and validation of a rule-based, regular expression model to quantify perioperative blood loss from EHRs. By providing estimates of median blood loss and interpatient variability across a wide range of procedures, the model addressed the current lack of standardised, data-driven definitions of excessive surgical bleeding. These findings offer a basis for defining pathological bleeding and may support future efforts to identify associated risk factors and inform perioperative planning and quality assessment.

The third study applied NLP-based deep learning models to detect and characterise major bleeding events in acutely hospitalised medical patients. By using unstructured clinical text, the models identified relevant bleeding events not captured by structured data alone. The analysis suggested that bleeding risk in this population was more closely related to physiological frailty and abnormal laboratory values than to anticoagulant use, indicating the need for broader approaches to bleeding risk assessment.

The final study focused on the design, implementation, and early clinical evaluation of an NLP system to extract thrombosis- and bleeding-related risk factors from unstructured clinical documentation in support of thrombophilia assessment. The system was successfully integrated into clinical workflows and improved the focus and efficiency of record review. However, it also highlighted key implementation challenges, including automation bias, interface usability, and the importance of clinician trust, which require careful consideration in future developments.

Collectively, these studies demonstrate how NLP methods, ranging from transparent rule-based systems to more complex deep learning approaches, can help address existing limitations in haemostatic risk assessment and clinical workflows. While further refinement, evaluation, and regulatory consideration are necessary, this work contributes to the growing field of data-driven tools aimed at supporting more consistent, informed, and patient-centred care in bleeding and thrombosis management.

References

- 1 Alnor AB, Gils C, Vinholt PJ. Venous thromboembolism risk in adults with hereditary thrombophilia: a systematic review and meta-analysis. *Ann Hematol*. 2024; **103**: 4285-94. 10.1007/s00277-024-05926-2.
- 2 Alnor AB, Lynggaard RB, Laursen MS, Vinholt PJ. Natural language processing for identifying major bleeding risk in hospitalised medical patients. *Comput Biol Med*. 2025; **190**: 110093. 10.1016/j.combiomed.2025.110093.
- 3 Denis CV, Lenting PJ. von Willebrand factor: at the crossroads of bleeding and thrombosis. *Int J Hematol*. 2012; **95**: 353-61. 10.1007/s12185-012-1041-x.
- 4 Shah A, Kerner V, Stanworth SJ, Agarwal S. Major haemorrhage: past, present and future. *Anaesthesia*. 2023; **78**: 93-104. 10.1111/anae.15866.
- 5 Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. *J Thromb Thrombolysis*. 2016; **41**: 3-14. 10.1007/s11239-015-1311-6.
- 6 Anderson FA, Jr., Spencer FA. Risk factors for venous thromboembolism. *Circulation*. 2003; **107**: I9-16. 10.1161/01.CIR.0000078469.07362.E6.
- 7 Campello E, Spiezia L, Adamo A, Simioni P. Thrombophilia, risk factors and prevention. *Expert Rev Hematol*. 2019; **12**: 147-58. 10.1080/17474086.2019.1583555.

- 8 Pastori D, Cormaci VM, Marucci S, Franchino G, Del Sole F, Capozza A, Fallarino A, Corso C, Valeriani E, Menichelli D, Pignatelli P. A Comprehensive Review of Risk Factors for Venous Thromboembolism: From Epidemiology to Pathophysiology. *Int J Mol Sci.* 2023; **24**. 10.3390/ijms24043169.
- 9 Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the S, Standardization Committee of the International Society on T, Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost.* 2005; **3**: 692-4. 10.1111/j.1538-7836.2005.01204.x.
- 10 Decousus H, Tapson VF, Bergmann JF, Chong BH, Froehlich JB, Kakkar AK, Merli GJ, Monreal M, Nakamura M, Pavanello R, Pini M, Piovela F, Spencer FA, Spyropoulos AC, Turpie AG, Zotz RB, Fitzgerald G, Anderson FA, Investigators I. Factors at admission associated with bleeding risk in medical patients: findings from the IMPROVE investigators. *Chest.* 2011; **139**: 69-79. 10.1378/chest.09-3081.
- 11 Darzi AJ, Karam SG, Spencer FA, Spyropoulos AC, Mbuagbaw L, Woller SC, Zakai NA, Streiff MB, Gould MK, Cushman M, Charide R, Etzeandia-Ikobaltzeta I, Germini F, Rigoni M, Agarwal A, Morsi RZ, Akl EA, Iorio A, Schunemann HJ. Risk models for VTE and bleeding in medical inpatients: systematic identification and expert assessment. *Blood Adv.* 2020; **4**: 2557-66. 10.1182/bloodadvances.2020001937.
- 12 Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med.* 1965; **58**: 295-300. 10.1177/003591576505800503.
- 13 Oger E, Botrel MA, Juchault C, Bouget J. Sensitivity and specificity of an algorithm based on medico-administrative data to identify hospitalized patients with major bleeding presenting to an emergency department. *BMC Med Res Methodol.* 2019; **19**: 194. 10.1186/s12874-019-0841-6.
- 14 Al-Ani F, Shariff S, Siqueira L, Seyam A, Lazo-Langner A. Identifying venous thromboembolism and major bleeding in emergency room discharges using administrative data. *Thromb Res.* 2015; **136**: 1195-8. 10.1016/j.thromres.2015.10.035.
- 15 Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Res Synth Methods.* 2017; **8**: 537-53. 10.1002/jrsm.1260.
- 16 Weingarz L, Schindewolf M, Schwonberg J, Hecking C, Wolf Z, Erbe M, Lindhoff-Last E, Linnemann B. Thrombophilia and risk of VTE recurrence according to the age at the time of first VTE manifestation. *Vasa.* 2015; **44**: 313-23. 10.1024/0301-1526/a000447.
- 17 Eppenberger D, Nilius H, Anagnostelis B, Huber CA, Nagler M. Current Knowledge on Factor V Leiden Mutation as a Risk Factor for Recurrent Venous Thromboembolism: A Systematic Review and Meta-Analysis. *Frontiers in Cardiovascular Medicine.* 2022; **9**. ARTN 883986. 10.3389/fcvm.2022.883986.
- 18 Federici EH, Al-Mondhry H. High risk of thrombosis recurrence in patients with homozygous and compound heterozygous factor V R506Q (Factor V Leiden) and prothrombin G20210A. *Thromb Res.* 2019; **182**: 75-8. 10.1016/j.thromres.2019.07.030.
- 19 Ho WK, Hankey GJ, Quinlan DJ, Eikelboom JW. Risk of recurrent venous thromboembolism in patients with common thrombophilia: a systematic review. *Arch Intern Med.* 2006; **166**: 729-36. 10.1001/archinte.166.7.729.
- 20 Marchiori A, Mosena L, Prins MH, Prandoni P. The risk of recurrent venous thromboembolism among heterozygous carriers of factor V Leiden or prothrombin G20210A mutation. A systematic review of prospective studies. *Haematologica.* 2007; **92**: 1107-14. 10.3324/haematol.10234.
- 21 Croles FN, Nasserinejad K, Duvekot JJ, Kruip MJ, Meijer K, Leebeek FW. Pregnancy, thrombophilia, and the risk of a first venous thrombosis: systematic review and bayesian meta-analysis. *BMJ.* 2017; **359**: j4452. 10.1136/bmj.j4452.
- 22 Coccolini F, Shander A, Ceresoli M, Moore E, Tian B, Parini D, Sartelli M, Sakakushev B, Doklestich K, Abu-Zidan F, Horer T, Shelat V, Hardcastle T, Bignami E, Kirkpatrick A, Weber D, Kryvoruchko I, Leppaniemi A, Tan E, Kessel B, Isik A, Cremonini C, Forfori F, Ghiadoni L, Chiarugi M, Ball C, Ottolino P, Hecker A, Mariani D, Melai E, Malbrain M, Agostini V, Podda M, Picetti E, Kluger Y, Rizoli S, Litvin A, Maier R, Beka SG, De Simone B, Bala M, Perez AM, Ordóñez C, Bodnaruk Z, Cui Y, Calatayud AP, de Angelis N, Amico

- F, Pikoulis E, Damaskos D, Coimbra R, Chirica M, Biffi WL, Catena F. Strategies to prevent blood loss and reduce transfusion in emergency general surgery, WSES-AAST consensus paper. *World J Emerg Surg*. 2024; **19**: 26. 10.1186/s13017-024-00554-7.
- 23 Kietai S, Ahmed A, Afshari A, Albaladejo P, Aldecoa C, Barauskas G, De Robertis E, Faraoni D, Filipescu DC, Fries D, Godier A, Haas T, Jacob M, Lance MD, Llau JV, Meier J, Molnar Z, Mora L, Rahe-Meyer N, Samama CM, Scarlatescu E, Schlimp C, Wikkelsso AJ, Zacharowski K. Management of severe peri-operative bleeding: Guidelines from the European Society of Anaesthesiology and Intensive Care: Second update 2022. *Eur J Anaesthesiol*. 2023; **40**: 226-304. 10.1097/EJA.0000000000001803.
- 24 Arachchillage DJ, Mackillop L, Chandratheva A, Motawani J, MacCallum P, Laffan M. Thrombophilia testing: A British Society for Haematology guideline. *Br J Haematol*. 2022; **198**: 443-58. 10.1111/bjh.18239.
- 25 Pedersen JS, Laursen MS, Soguero-Ruiz C, Savarimuthu TR, Hansen RS, Vinholt PJ. Domain over size: Clinical ELECTRA surpasses general BERT for bleeding site classification in the free text of electronic health records. *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*: IEEE, 2022, 1-4.
- 26 Croles FN, Borjas-Howard J, Nasserinejad K, Leebeek FWG, Meijer K. Risk of Venous Thrombosis in Antithrombin Deficiency: A Systematic Review and Bayesian Meta-analysis. *Semin Thromb Hemost*. 2018; **44**: 315-26. 10.1055/s-0038-1625983.
- 27 Di Minno MN, Ambrosino P, Ageno W, Rosendaal F, Di Minno G, Dentali F. Natural anticoagulants deficiency and the risk of venous thromboembolism: a meta-analysis of observational studies. *Thromb Res*. 2015; **135**: 923-32. 10.1016/j.thromres.2015.03.010.
- 28 Middeldorp S, Nieuwlaet R, Baumann Kreuziger L, Coppens M, Houghton DE, James AH, Lang E, Moll S, Myers T, Bhatt M, Chai-Adisaksoha C, Colunga-Lozano LE, Karam SG, Zhang Y, Wiercioch W, Schunemann HJ, Iorio A. American Society of Hematology 2023 Guidelines for Management of Venous Thromboembolism: Thrombophilia Testing. *Blood Adv*. 2023. 10.1182/bloodadvances.2023010177.
- 29 Marcucci M, Etzeandía-Ikobaltzeta I, Yang S, Germini F, Gupta S, Agarwal A, Ventresca M, Tang S, Morgano GP, Wang M, Ahmed MM, Neumann I, Izcovich A, Criniti J, Popoff F, Devereaux PJ, Dahm P, Anderson D, Lavikainen LI, Tikkinen KAO, Guyatt GH, Schunemann HJ, Violette PD. Benefits and harms of direct oral anticoagulation and low molecular weight heparin for thromboprophylaxis in patients undergoing non-cardiac surgery: systematic review and network meta-analysis of randomised trials. *BMJ*. 2022; **376**: e066785. 10.1136/bmj-2021-066785.
- 30 Salenger R, Arora RC, Bracey A, D'Oria M, Engelman DT, Evans C, Grant MC, Gunaydin S, Morton V, Ozawa S, Patel PA, Raphael J, Rosengart TK, Shore-Lesserson L, Tibi P, Shander A. Cardiac Surgical Bleeding, Transfusion, and Quality Metrics: Joint Consensus Statement by the Enhanced Recovery After Surgery Cardiac Society and Society for the Advancement of Patient Blood Management. *Ann Thorac Surg*. 2025; **119**: 280-95. 10.1016/j.athoracsur.2024.06.039.
- 31 Mullen L, Stables R. Validation of HES coding for the detection of major bleeding events: insights from the ROBOT-ACS study. *BMC Med Res Methodol*. 2025; **25**: 42. 10.1186/s12874-025-02503-7.
- 32 Nopp S, Ay C. Bleeding Risk Assessment in Patients with Venous Thromboembolism. *Hamostaseologie*. 2021; **41**: 267-74. 10.1055/a-1339-9987.
- 33 Singh M, Kumar A, Khanna NN, Laird JR, Nicolaides A, Faa G, Johri AM, Mantella LE, Fernandes JFE, Teji JS, Singh N, Fouda MM, Singh R, Sharma A, Kitis G, Rathore V, Singh IM, Tadepalli K, Al-Maini M, Isenovic ER, Chaturvedi S, Garg D, Paraskevas KI, Mikhailidis DP, Viswanathan V, Kalra MK, Ruzsa Z, Saba L, Laine AF, Bhatt DL, Suri JS. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. *EClinicalMedicine*. 2024; **73**: 102660. 10.1016/j.eclinm.2024.102660.
- 34 Wang J, Wang Z, Chen M, Xiao Y, Chen S, Wu L, Yao L, Jiang X, Li J, Xu M, Lin M, Zhu Y, Luo R, Zhang C, Li X, Yu H. An interpretable artificial intelligence system for detecting risk factors of gastroesophageal variceal bleeding. *NPJ Digit Med*. 2022; **5**: 183. 10.1038/s41746-022-00729-z.

- 35 Joos C, Lawrence K, Jones AE, Johnson SA, Witt DM. Accuracy of ICD-10 codes for identifying hospitalizations for acute anticoagulation therapy-related bleeding events. *Thromb Res*. 2019; **181**: 71-6. 10.1016/j.thromres.2019.07.021.
- 36 Shehab N, Ziemba R, Campbell KN, Geller AI, Moro RN, Gage BF, Budnitz DS, Yang TH. Assessment of ICD-10-CM code assignment validity for case finding of outpatient anticoagulant-related bleeding among Medicare beneficiaries. *Pharmacoepidemiol Drug Saf*. 2019; **28**: 951-64. 10.1002/pds.4783.
- 37 Subat YW, Rayes H, Hanson AC, Johnson MQ, Schulte PJ, Evans K, Weister T, Trivedi V, Gajic O, Warner MA. Risk of major bleeding associated with aspirin use in non-surgical critically ill patients receiving therapeutic anticoagulation. *J Crit Care*. 2020; **58**: 34-40. 10.1016/j.jcrc.2020.04.003.
- 38 Castellucci LA, Le Gal G, Rodger MA, Carrier M. Major bleeding during secondary prevention of venous thromboembolism in patients who have completed anticoagulation: a systematic review and meta-analysis. *J Thromb Haemost*. 2014; **12**: 344-8. 10.1111/jth.12501.
- 39 Recsky C, Rush KL, MacPhee M, Stowe M, Blackburn L, Muniak A, Currie LM. Clinical Informatics Team Members' Perspectives on Health Information Technology Safety After Experiential Learning and Safety Process Development: Qualitative Descriptive Study. *JMIR Form Res*. 2024; **8**: e53302. 10.2196/53302.
- 40 Kucking F, Hubner U, Przysucha M, Hannemann N, Kutza JO, Moelleken M, Erfurt-Berge C, Dissemond J, Babitsch B, Busch D. Automation Bias in AI-Decision Support: Results from an Empirical Study. *Stud Health Technol Inform*. 2024; **317**: 298-304. 10.3233/SHTI240871.
- 41 Wang DY, Ding J, Sun AL, Liu SG, Jiang D, Li N, Yu JK. Artificial intelligence suppression as a strategy to mitigate artificial intelligence automation bias. *J Am Med Inform Assoc*. 2023; **30**: 1684-92. 10.1093/jamia/ocad118.
- 42 Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of Electronic Health Record Use With Physician Fatigue and Efficiency. *JAMA Netw Open*. 2020; **3**: e207385. 10.1001/jamanetworkopen.2020.7385.
- 43 Roman LC, Ancker JS, Johnson SB, Senathirajah Y. Navigation in the electronic health record: A review of the safety and usability literature. *J Biomed Inform*. 2017; **67**: 69-79. 10.1016/j.jbi.2017.01.005.
- 44 Trinkley KE, Maw AM, Torres CH, Huebschmann AG, Glasgow RE. Applying Implementation Science to Advance Electronic Health Record-Driven Learning Health Systems: Case Studies, Challenges, and Recommendations. *J Med Internet Res*. 2024; **26**: e55472. 10.2196/55472.
- 45 Kamel Rahimi A, Pienaar O, Ghadimi M, Canfell OJ, Pole JD, Shrapnel S, van der Vegt AH, Sullivan C. Implementing AI in Hospitals to Achieve a Learning Health System: Systematic Review of Current Enablers and Barriers. *J Med Internet Res*. 2024; **26**: e49655. 10.2196/49655.
- 46 Houssein EH, Mohamed RE, Ali AA. Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques. *Sci Rep*. 2023; **13**: 7173. 10.1038/s41598-023-34294-6.
- 47 Afshar M, Adelaine S, Resnik F, Mundt MP, Long J, Leaf M, Ampian T, Wills GJ, Schnapp B, Chao M, Brown R, Joyce C, Sharma B, Dligach D, Burnside ES, Mahoney J, Churpek MM, Patterson BW, Liao F. Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults. *JMIR Med Inform*. 2023; **11**: e44977. 10.2196/44977.
- 48 Lam BD, Chrysafi P, Chiasakul T, Khosla H, Karagkouni D, McNichol M, Adamski A, Reyes N, Abe K, Mantha S, Vlachos IS, Zwicker JJ, Patell R. Machine learning natural language processing for identifying venous thromboembolism: systematic review and meta-analysis. *Blood Adv*. 2024; **8**: 2991-3000. 10.1182/bloodadvances.2023012200.
- 49 Johnson SA, Signor EA, Lappe KL, Shi J, Jenkins SL, Wikstrom SW, Kroencke RD, Hallowell D, Jones AE, Witt DM. A comparison of natural language processing to ICD-10 codes for identification and characterization of pulmonary embolism. *Thromb Res*. 2021; **203**: 190-5. 10.1016/j.thromres.2021.04.020.

- 50 Shi J, Hurdle JF, Johnson SA, Ferraro JP, Skarda DE, Finlayson SRG, Samore MH, Bucher BT. Natural language processing for the surveillance of postoperative venous thromboembolism. *Surgery*. 2021; **170**: 1175-82. 10.1016/j.surg.2021.04.027.
- 51 Woller B, Daw A, Aston V, Lloyd J, Snow G, Stevens SM, Woller SC, Jones P, Bledsoe J. Natural Language Processing Performance for the Identification of Venous Thromboembolism in an Integrated Healthcare System. *Clin Appl Thromb Hemost*. 2021; **27**: 10760296211013108. 10.1177/10760296211013108.
- 52 Scott DM, Abud A, Amor M, Saemi A, Baumann Kreuziger L, Houghton D. Natural Language Processing tool accurately identifies acute venous thromboembolism. *Thromb Res*. 2023; **229**: 252-4. 10.1016/j.thromres.2023.08.007.
- 53 Jin ZG, Zhang H, Tai MH, Yang Y, Yao Y, Guo YT. Natural Language Processing in a Clinical Decision Support System for the Identification of Venous Thromboembolism: Algorithm Development and Validation. *J Med Internet Res*. 2023; **25**: e43153. 10.2196/43153.
- 54 Shung D, Tsay C, Laine L, Chang D, Li F, Thomas P, Partridge C, Simonov M, Hsiao A, Tay JK, Taylor A. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. *J Gastroenterol Hepatol*. 2021; **36**: 1590-7. 10.1111/jgh.15313.
- 55 Taggart M, Chapman WW, Steinberg BA, Ruckel S, Pregoner-Wenzler A, Du Y, Ferraro J, Bucher BT, Lloyd-Jones DM, Rondina MT, Shah RU. Comparison of 2 Natural Language Processing Methods for Identification of Bleeding Among Critically Ill Patients. *JAMA Netw Open*. 2018; **1**: e183451. 10.1001/jamanetworkopen.2018.3451.
- 56 Munoz Martin AJ, Lecumberri R, Souto JC, Obispo B, Sanchez A, Aparicio J, Aguayo C, Gutierrez D, Garcia Palomo A, Benavent D, Taberna M, Vinuela-Beneitez MC, Arumi D, Hernandez-Presa MA. Prediction model for major bleeding in anticoagulated patients with cancer-associated venous thromboembolism using machine learning and natural language processing. *Clin Transl Oncol*. 2025; **27**: 1816-25. 10.1007/s12094-024-03586-2.
- 57 Shung DL, Chan CE, You K, Nakamura S, Saarinen T, Zheng NS, Simonov M, Li DK, Tsay C, Kawamura Y, Shen M, Hsiao A, Sekhon JS, Laine L. Validation of an Electronic Health Record-Based Machine Learning Model Compared With Clinical Risk Scores for Gastrointestinal Bleeding. *Gastroenterology*. 2024; **167**: 1198-212. 10.1053/j.gastro.2024.06.030.
- 58 Hsu E, Bako AT, Potter T, Pan AP, Britz GW, Tannous J, Vahidy FS. Extraction of Radiological Characteristics From Free-Text Imaging Reports Using Natural Language Processing Among Patients With Ischemic and Hemorrhagic Stroke: Algorithm Development and Validation. *JMIR AI*. 2023; **2**: e42884. 10.2196/42884.
- 59 Girouard MP, Chang AJ, Liang Y, Hamilton SA, Bhatt AS, Svetlichnaya J, Fitzpatrick JK, Carey ECB, Avula HR, Adatya S, Lee KK, Solomon MD, Parikh RV, Go AS, Ambrosy AP. Clinical and research applications of natural language processing for heart failure. *Heart Fail Rev*. 2025; **30**: 407-15. 10.1007/s10741-024-10472-0.
- 60 Yan MY, Gustad LT, Nytro O. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. *J Am Med Inform Assoc*. 2022; **29**: 559-75. 10.1093/jamia/ocab236.
- 61 Wei J, Hu T, Dai J, Wang Z, Han P, Huang W. Research on named entity recognition of adverse drug reactions based on NLP and deep learning. *Front Pharmacol*. 2023; **14**: 1121796. 10.3389/fphar.2023.1121796.
- 62 Khan F, Triteschler T, Kahn SR, Rodger MA. Venous thromboembolism. *Lancet*. 2021; **398**: 64-77. 10.1016/S0140-6736(20)32658-1.
- 63 Pisters R, Lane DA, Nieuwlaet R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest*. 2010; **138**: 1093-100. 10.1378/chest.10-0134.

- 64 Mascarenhas M, Mendes F, Martins M, Ribeiro T, Afonso J, Cardoso P, Ferreira J, Fonseca J, Macedo G. Explainable AI in Digestive Healthcare and Gastrointestinal Endoscopy. *J Clin Med*. 2025; **14**. 10.3390/jcm14020549.
- 65 Ce M, Ibba S, Cellina M, Tancredi C, Fantesini A, Fazzini D, Fortunati A, Perazzo C, Presta R, Montanari R, Forzenigo L, Carrafiello G, Papa S, Ali M. Radiologists' perceptions on AI integration: An in-depth survey study. *Eur J Radiol*. 2024; **177**: 111590. 10.1016/j.ejrad.2024.111590.
- 66 Goh WW, Chia KY, Cheung MF, Kee KM, Lwin MO, Schulz PJ, Chen M, Wu K, Ng SS, Lui R, Ang TL, Yeoh KG, Chiu HM, Wu DC, Sung JJ. Risk Perception, Acceptance, and Trust of Using AI in Gastroenterology Practice in the Asia-Pacific Region: Web-Based Survey Study. *JMIR AI*. 2024; **3**: e50525. 10.2196/50525.
- 67 Lim SS, Phan TD, Law M, Goh GS, Moriarty HK, Lukies MW, Joseph T, Clements W. Non-radiologist perception of the use of artificial intelligence (AI) in diagnostic medical imaging reports. *J Med Imaging Radiat Oncol*. 2022; **66**: 1029-34. 10.1111/1754-9485.13388.
- 68 Hah H, Goldin DS. How Clinicians Perceive Artificial Intelligence-Assisted Technologies in Diagnostic Decision Making: Mixed Methods Approach. *J Med Internet Res*. 2021; **23**: e33540. 10.2196/33540.
- 69 Correia Pinheiro L, Arlett P, Roes K, Musuamba Tshinanu F, Westman G, Frias Z, Hamann H, Berenguer Jornet J, Khan I, Larsen J, Broich K, Cooke E. Artificial Intelligence in European Medicines Regulation: From Vision to Action. Harnessing the Capabilities of Artificial Intelligence for the Benefit of Public and Animal Health. *Clin Pharmacol Ther*. 2025; **117**: 335-6. 10.1002/cpt.3494.