ORIGINAL RESEARCH

Incidence and Risk Factors of Lower Limb Deep Vein Thrombosis in Psychiatric Inpatients by Applying Machine Learning to Electronic Health Records: A Retrospective Cohort Study

Liang Xu ⓘD, Miao Da ⓘD

Department of Psychiatry, Huzhou Third Municipal Hospital, the Affiliated Hospital of Huzhou University, Huzhou, Zhejiang, People's Republic of China

Correspondence: Miao Da, Department of Psychiatry, Huzhou Third Municipal Hospital, the Affiliated Hospital of Huzhou University, 2088 East Tiaoxi Road, Huzhou, Zhejiang, People's Republic of China, Tel +860572 2290561, Email dm2315891089@163.com

Background: Psychiatric inpatients face an increased risk of deep vein thrombosis (DVT) due to their psychiatric conditions and pharmacological treatments. However, research focusing on this population remains limited.
Methods: This study analyzed 17,434 psychiatric inpatients at Huzhou Third Municipal Hospital, incorporating data on demographics, psychiatric diagnoses, physical illnesses, laboratory results, and medication use. Predictive models for DVT were developed using logistic regression, random forest, support vector machine (SVM), and XGBoost (Extreme Gradient Boosting). Feature importance was assessed using the random forest model.
Results: The DVT incidence among psychiatric inpatients was 1.6%. Predictive model performance, measured by the area under the curve (AUC), showed logistic regression (0.900), random forest (0.885), SVM (0.890), and XGBoost (0.889) performed well. Logistic regression and random forest models exhibited optimal overall performance, while XGBoost excelled in recall. Significant predictors of DVT included elevated D-dimer levels, age, Alzheimer's disease, and Madopar use.
Conclusion: Psychiatric inpatients require vigilance for DVT risk, with factors like D-dimer levels and age serving as critical indicators. Machine learning models effectively predict DVT risk, enabling early detection and personalized prevention strategies in clinical practice.
Keywords: psychiatric inpatients, deep vein thrombosis, machine learning, risk factors, predictive modelling

Introduction

Recent years have seen a steady rise in the number of psychiatric inpatients, a trend linked to the aging population and the increasing incidence of mental illness.[1] While psychiatric inpatient care primarily addresses psychiatric and behavioral symptoms, the physical health concerns associated with these patients, particularly the risk of deep vein thrombosis (DVT), have often been overlooked. DVT is a common and potentially serious complication in hospitalized patients, especially those who are bedridden or have limited mobility. In addition to raising healthcare costs during hospitalization, DVT can lead to severe complications, such as pulmonary embolism.[2]

The thrombotic risk for psychiatric inpatients is often exacerbated by the nature of their condition. For instance, long-term use of antipsychotic medications can lead to metabolic syndrome, while suicidal and self-harming behaviors, along with overall reduced physical activity, further heighten the risk of thrombosis.[3–5] Additionally, psychiatric patients frequently have comorbid somatic conditions such as hypertension and diabetes, which are also associated with an increased risk of thrombosis.[6]

Existing literature highlights that factors such as D-dimer levels, age, and thrombotic history are strongly associated with the occurrence of DVT. However, most studies have primarily examined general medical and surgical populations,

with a notable lack of research focused on psychiatric inpatients.[7–9] Given the complex array of factors contributing to DVT in this population and the intricate interrelationships among these factors, conventional statistical analysis methods may face limitations in adequately addressing these multifaceted interactions.

To address the limitations of traditional statistical methods and offer a more thorough analysis of DVT risk factors in psychiatric inpatients, this study adopts a machine learning approach. As a data-driven analytical technique, machine learning enables the automated detection of patterns and trends within complex datasets.[10] Its use in medical research has been expanding, particularly for multidimensional and high-dimensional data.[11] Compared to traditional statistical methods, machine learning is superior in handling non-linear relationships and uncovering intricate interactions between multiple variables.[12]

The objective of this study is a notable lack in the field of thrombosis in psychiatric inpatients through the application of machine learning methods, thereby contributing to the advancement of clinical practice. This study specifically focuses on the incidence and risk factors of lower extremity DVT among psychiatric inpatients, rather than the broader category of venous thromboembolic disease which also includes PE. The decision to limit our study to lower extremity DVT was based on its high incidence in this population and the relative ease of diagnosis compared to PE, which often requires additional imaging techniques such as CT pulmonary angiography. The findings contribute to the efficiency of early prevention and intervention of DVT, thereby improving the overall prognosis and quality of life in this population.

Methods
Study Population
A total of 17,434 patients admitted to the psychiatric department of Huzhou Third Municipal Hospital between July 2021 and May 2024 were included in the study. The inclusion criteria were as follows: patients who were admitted to the psychiatric department during the specified period, with medical records that provided comprehensive details on their medical history, diagnostic information, laboratory test results, and medication during hospitalization. The principal diagnosis for each patient was a mental or behavioral disorder, as classified by the International Classification of Diseases, 10th edition (ICD-10), within the range of codes F00-F99. For patients with multiple hospitalizations, data from the first hospitalization was used. Exclusion criteria included: Patients who died during hospitalization. Medical records that were incomplete or lacked essential data for comprehensive analysis. Patients with an insufficient length of stay to collect necessary data. Patients with pre-existing deep vein thrombosis at the time of admission. Additional data from multiple hospitalizations beyond the first hospitalization for patients with multiple admissions.

Data Collection
The researcher accessed patient data from the hospital's electronic medical record system, which included the following information:

Demographic Information
Age and gender.

Medical History
Length of hospitalization and comorbid physical conditions, such as hypertension and diabetes.

Diagnostic Information
Patient diagnoses of mental disorders were classified according to the International Classification of Diseases, 10th Revision (ICD-10). Major diagnostic categories included Alzheimer's disease, dementia, schizophrenia, depressive disorder, and bipolar disorder.

Laboratory Test Results
Data from 75 blood tests were collected, including hemoglobin, white blood cell count, platelet count, alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatinine, urea nitrogen, total cholesterol, low-density lipoprotein (LDL), blood glucose, D-dimer, and troponin T, among others.

Drug Use

The study reviewed the use of 76 commonly prescribed drugs in psychiatry, including antipsychotics, antidepressants, anti-anxiety medications, anti-dementia drugs, and medications used for comorbid somatic conditions.

DVT Diagnosis

The primary outcome of interest was the diagnosis of DVT occurring during hospitalization. DVT was diagnosed based on color Doppler ultrasound of both lower limb veins, which was the established diagnostic criterion. The diagnostic criteria included an inability to compress and occlude the vein lumen, as well as the presence of hypoechoic or echogenic areas with minimal or no blood flow signal within the vein.[13] Patients were categorized into two groups: those who developed DVT during their hospitalization and those who did not.

Data Preprocessing

Duplicate features were removed based on expert recommendations. Variables or research objects with a missing data rate exceeding 10% were excluded from the analysis. For the remaining missing data, the K-nearest neighbor (KNN) method was employed to impute values. After consolidating and imputing the data, preprocessing steps were undertaken to optimize machine learning computations. Continuous variables with near-zero variance were eliminated, and the data were normalized by standardizing them to have a mean of zero and a standard deviation of one. This preprocessing ensured that all indicators were on a comparable scale, facilitating a comprehensive comparative analysis. Categorical variables were converted into factor variables.

Model Development and Validation

The model was developed using four prevalent machine learning algorithms: logistic regression, random forest, support vector machine (SVM), and XGBoost. To mitigate the data imbalance, the under-sampling technique was applied, which reduced the number of instances in the majority class to enhance the representation of the minority class.[14] The dataset was randomly split into training and validation sets with a 7:3 ratio. The training set was used to build the model, while the validation set was used for evaluation.

Feature selection was performed using LASSO (Least Absolute Shrinkage and Selection Operator) regression. LASSO regression, through L1 regularization, automatically selected features and reduced the number of independent variables in the model. The feature variables selected by LASSO regression were incorporated into the predictive model. To construct the model, ten-fold cross-validation (CV) was employed, wherein the training set was divided into ten equal parts. Each part served sequentially as the validation set, while the remaining nine parts were used for training. The model's final performance metrics were derived from averaging the results of these ten iterations, providing a robust assessment of its generalization capabilities.

Model parameters were optimized through grid search. Performance was evaluated using metrics such as accuracy, precision, recall, F1 score, and AUC (area under the curve). ROC curves for each model were plotted, with AUC values interpreted as follows: $0.5 \leq AUC < 0.7$ indicated poor prediction, $0.7 \leq AUC < 0.9$ indicated moderate prediction, and $AUC \geq 0.9$ indicated very good prediction.

Feature importance was assessed using the Random Forest algorithm, which improved model performance and stability by aggregating predictions from multiple decision trees. The importance of each feature was determined by calculating its average score across these trees, indicating its contribution to model performance. The study's methodology was illustrated in Figure 1.

Statistical methods

The statistical analysis and modeling were performed using IBM SPSS version 26.0, Python 3.8, Anaconda distribution 23.1.0, and the Spyder development environment 5.4.1. For normally distributed data, results were presented as means ± standard deviations, and comparisons between groups were conducted using t-tests. For non-normally distributed data, results were reported as medians with interquartile ranges [M (Q1, Q3)], and the Mann–Whitney U-test was used for
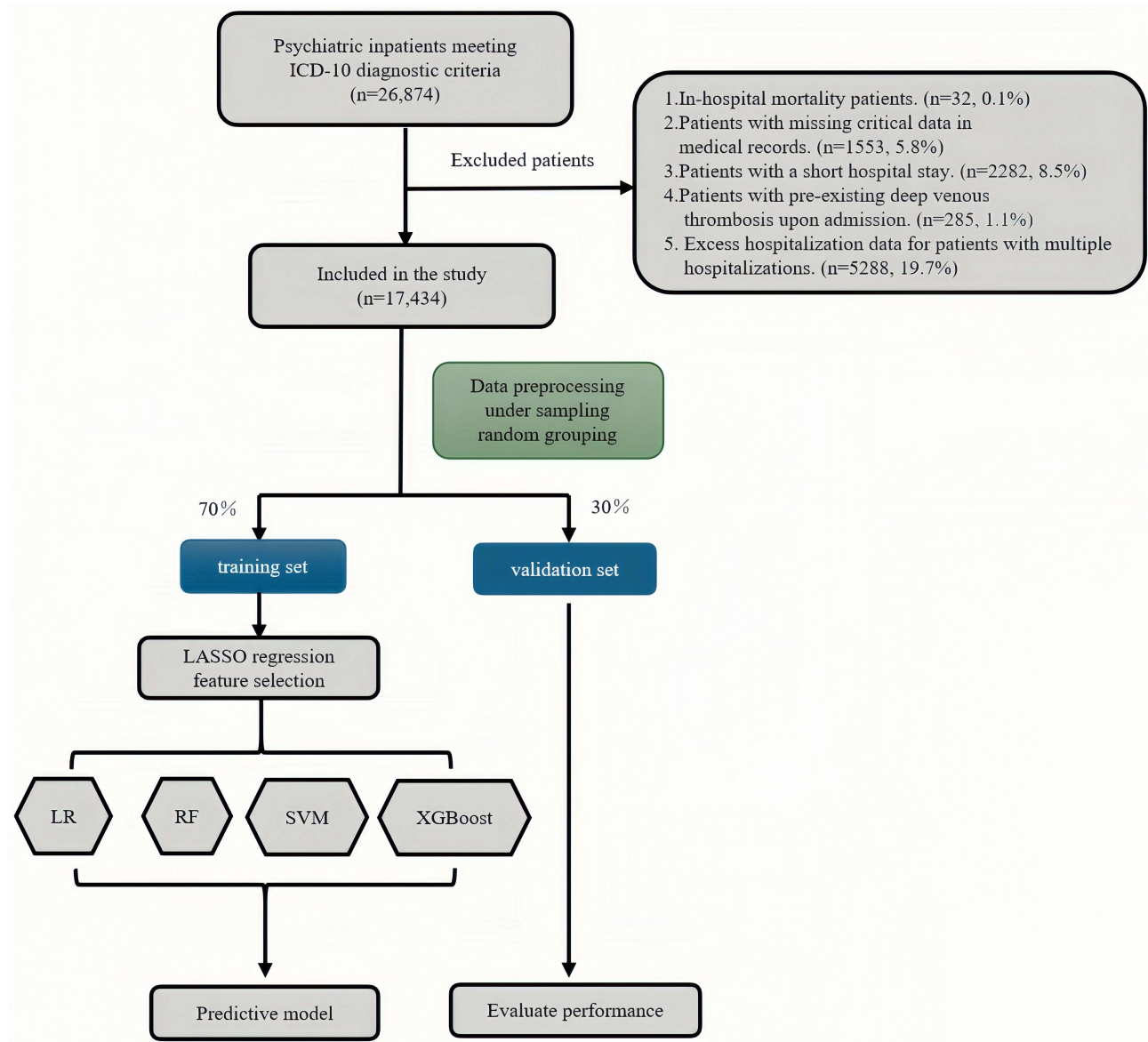
Figure 1 Study flowchart. This figure outlined the process of patient selection, data preprocessing, and model evaluation for a study involving psychiatric inpatients who met the ICD-10 diagnostic criteria. The flowchart detailed the steps taken to refine the dataset and the predictive models used for analysis.
Abbreviations: ICD-10, International Classification of Diseases, Tenth Revision; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; RF, Random Forest; SVM, Support Vector Machine; XGBoost, Extreme Gradient Boosting.

group comparisons. Categorical data were expressed as frequencies and percentages, with group comparisons analyzed using the χ^2-test or Fisher's exact test. A p-value of less than 0.05 was considered indicative of statistical significance.

Results
Clinical Features of Patients
A total of 17,434 inpatients with mental disorders were selected for the study. These participants were randomly divided into two groups: a training set consisting of 70% of the total sample (n = 13,947) and a validation set comprising the remaining 30% (n = 3,487). The mean age of the patients was 54.0 years (range: 35.0 to 65.0). Among them, 5,743 (32.94%) were male. The average length of hospital stay was 14.0 days (range: 10.0 to 21.0), and the mean D-dimer level was 0.26 mg/L (range: 0.15 to 0.51). Key characteristics of the data are detailed in Table 1.

Table 1 Comparison of General Information Between Thrombus and Non-Thrombus Groups

Characteristics	Patients Enrolled (n=17434)	Thrombus Groups (n=283)	Non-Thrombus Groups (n=17151)	p	Estimate (95% Confidence Interval)
Sex [Number of cases (%), male]	5743 (32.94%)	84 (29.68%)	5659 (33.00%)	0.266	3.32% (−1.1–7.7%)
Age [M (Q1, Q3), years]	54.0 (35.0, 65.0)	72.0 (65.0, 77.0)	53.0 (35.0, 65.0)	0.000	19.0 (18.0–20.0)
Days of hospitalisation [M (Q1, Q3), days]	14.0 (10.0, 21.0)	17.0 (11.0, 28.0)	14.0 (10.0, 21.0)	0.000	3.0 (2.0–4.0)
Alanine aminotransferase [M (Q1, Q3), U/L]	17.0 (11.8, 26.8)	16.7 (12.1, 24.25)	17.0 (11.8, 26.8)	0.258	−0.3 (−1.0–0.5)
D-dimer [M (Q1, Q3), mg/L]	0.26 (0.15, 0.51)	1.83 (0.63, 3.8)	0.26 (0.15, 0.49)	0.000	1.57 (1.30–1.84)
Prothrombin time [M (Q1, Q3), seconds]	11.2 (10.7, 11.7)	11.4 (10.9, 11.9)	11.2 (10.7, 11.7)	0.000	0.2 (0.1–0.3)
Urea nitrogen [M(Q1, Q30), mmol/L]	4.5 (3.65, 5.52)	5.33 (4.3, 7.025)	4.49 (3.64, 5.5)	0.000	0.83 (0.65–1.01)
Uric acid [M(Q1, Q3), μmol/L]	298.6 (245.3, 365.7)	283.8 (233.55, 362.65)	298.9 (245.6, 365.7)	0.035	−15.1 (−30.2- −0.1)
Thyrotropin [M(Q1, Q3), μIU/mL]	1.75 (1.18, 2.58)	1.75 (1.125, 2.6)	1.75 (1.18, 2.58)	0.924	No significant difference
Triglyceride [M(Q1, Q3), mmol/L]	1.26 (0.9, 1.87)	1.18 (0.89, 1.7)	1.26 (0.9, 1.87)	0.202	−0.08 (−0.2–0.04)
White blood cell count [M (Q1, Q30, 10^9/L]	5.7 (4.67, 6.99)	5.68 (4.835, 7.32)	5.7 (4.67, 6.98)	0.084	$−0.02 \times 10^9$/L (−0.1–0.06)
Albumin [M (Q1,Q3), g/L]	40.7 (38.3, 43.3)	39.3 (36.05, 41.9)	40.7 (38.3, 43.4)	0.000	−1.4 (−2.0 - −0.8)
Phosphorus [M(Q1, Q3), mmol/L]	1.07 (0.95, 1.19)	0.97 (0.86, 1.1)	1.07 (0.95, 1.19)	0.000	−0.10 (−0.14 - −0.06)
Creatinine [M(Q1, Q3), μmol/L]	61.2 (52.9, 72.3)	64.9 (54.6, 81.6)	61.2 (52.8, 72.1)	0.000	3.7 (2.1–5.3)
Choline esterase [M (Q1, Q3), U/L]	8079.0 (6889.25, 9362.0)	7114.0 (6016.5, 8414.0)	8080.0 (6911.0, 9373.0)	0.000	−965.0 (−1500 - −430)
Glucose [M(Q1, Q3), mmol/L]	4.98 (4.53, 5.86)	5.8 (4.835, 7.0)	4.97 (4.53, 5.84)	0.000	0.82 (0.5–1.1)
Haemoglobin [M(Q1, Q3), g/L]	129.0 (120.0, 140.0)	122.0 (112.0, 132.0)	129.0 (120.0, 140.0)	0.000	−7.0 (−10.0 - −4.0)
C-reactive protein [M(Q1, Q3), mg/L]	1.12 (0.43, 3.08)	2.23 (0.85, 7.205)	1.12 (0.43, 3.04)	0.000	1.11 (0.4–1.8)
Potassium [M(Q1,Q3), mmol/L]	3.82 (3.59, 4.03)	3.78 (3.455, 4.0)	3.82 (3.59, 4.03)	0.003	−0.04 (−0.1–0.0)
Follicle stimulating hormone [M(Q1,Q3), mIU/mL]	9.28 (4.28, 51.24)	45.41 (15.395, 62.56)	9.28 (4.24, 50.805)	0.000	36.13 (28.0–44.0)

(Continued)

Table 1 (Continued).

Characteristics	Patients Enrolled (n=17434)	Thrombus Groups (n=283)	Non-Thrombus Groups (n=17151)	p	Estimate (95% Confidence Interval)
Lactogen [M(Q1,Q3), ng/mL]	284.0 (145.0, 467.75)	284.0 (172.5, 425.5)	284.0 (144.0, 468.0)	0.423	No significant difference
Myoglobin [M(Q1,Q3), ng/mL]	21.0 (21.0, 30.21)	27.23 (21.0, 57.3)	21.0 (21.0, 29.96)	0.000	6.23 (3.0–9.5)
Troponin T M(Q1,Q3), pg/mL]	5.07 (3.6, 7.12)	9.1 (6.045, 16.945)	5.07 (3.58, 7.02)	0.000	4.03 (2.1–6.0)
Hypertension [Number of cases (%)]	4723 (27.09%)	143 (50.53%)	4580 (26.70%)	0.000	23.44% (17.0–30.0%)
Diabetes mellitus [Number of cases (%)]	1624 (9.32%)	43 (15.19%)	1581 (9.22%)	0.001	5.87% (3.0–9.0%)
Schizophrenia [Number of cases (%)]	2141 (12.28%)	18 (6.36%)	2123 (12.38%)	0.003	-5.92%(-9.1-2.7%)
Alzheimer's disease [Number of cases (%)]	913 (5.24%)	55 (19.43%)	858 (5.00%)	0.000	14.43%(9.5-19.5%)
Bipolar disorder [Number of cases (%)]	1717 (9.85%)	18 (6.36%)	1699 (9.91%)	0.059	-3.49%(-6.8- -0.2%)
Olanzapine [Number of cases (%)]	5230 (30.00%)	103 (36.40%)	5127 (29.89%)	0.021	6.51% (2.0-11.0%)
Clozapine [Number of cases (%)]	836 (4.80%)	7 (2.47%)	829 (4.83%)	0.089	-2.33%(-5.3 - 0.6%)
Madopar [Number of cases (%)]	199 (1.14%)	18 (6.36%)	181 (1.06%)	0.000	5.30%(3.1 - 7.5%)

The Incidence of Lower Limb Deep Vein Thrombosis

The study findings revealed that 283 patients developed lower extremity DVT, yielding a incidence rate of 1.6%. The incidence of lower extremity DVT varied across different subpopulations. Specifically, Figure 2 showed a incidence of 9.05% in patients administered Madopar and 6.02% in those with Alzheimer's disease. Furthermore, the incidence of lower extremity DVT increased with age, as illustrated in Figure 3.

Feature Selection

This study included a total of 163 features, encompassing demographic information, psychiatric diagnoses, concomitant physical illnesses, laboratory findings, and substance use. These features were analyzed using LASSO regression, resulting in a final model that retained 16 key factors: D-dimer, age, troponin T, urea nitrogen, Alzheimer's disease, red blood cell count, myoglobin, follicle-stimulating hormone, phosphorus, hemoglobin, Madopar, cholinesterase, glucose, albumin, C-reactive protein (CRP), and creatinine. The indicators evaluated at $\lambda = 1$ standard error (1se) were used as modeling indicators, with the results detailed in Table 2. Indicators with coefficients of 0 at $\lambda = $ 1se were excluded from the table The selected features demonstrated high importance within the model and provided optimal performance with the minimal number of independent variables.
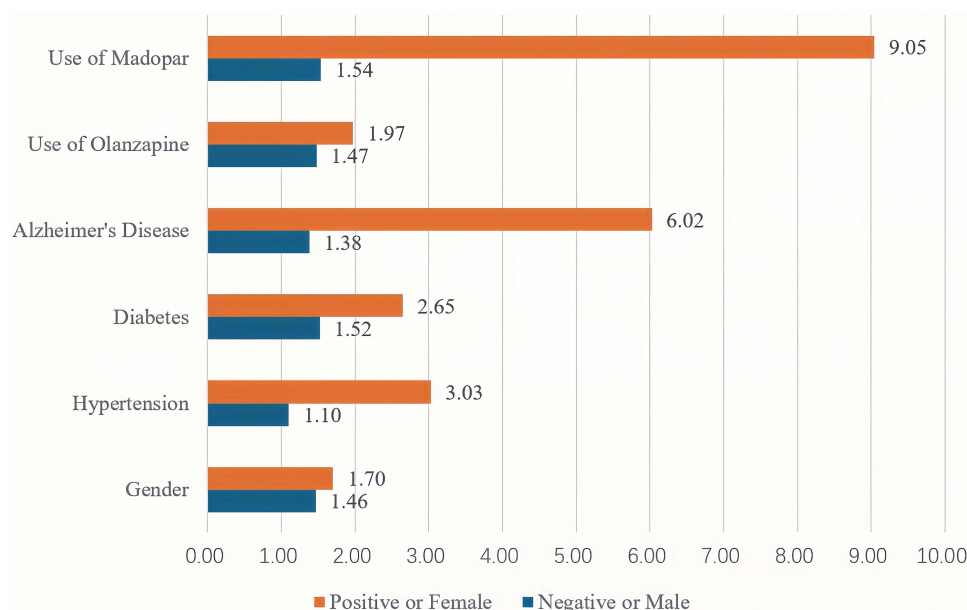
Figure 2 The incidence of Lower Limb Deep Vein Thrombosis (DVT) by Various Factors. This figure presented the incidence of lower limb deep vein thrombosis (DVT) as a percentage, categorized by different factors that could influence the occurrence of DVT. The graph displayed the rates for various conditions and characteristics, highlighting the relative risk associated with each factor. The x-axis listed the factors that are associated with DVT incidence, including the use of specific medications (eg, Madopar, Olanzapine), medical conditions (eg, Alzheimer's Disease, Diabetes, Hypertension), and gender (positive for female, negative for male). The y-axis represented the incidence rates of DVT, ranging from 0.00% to 10.00%.
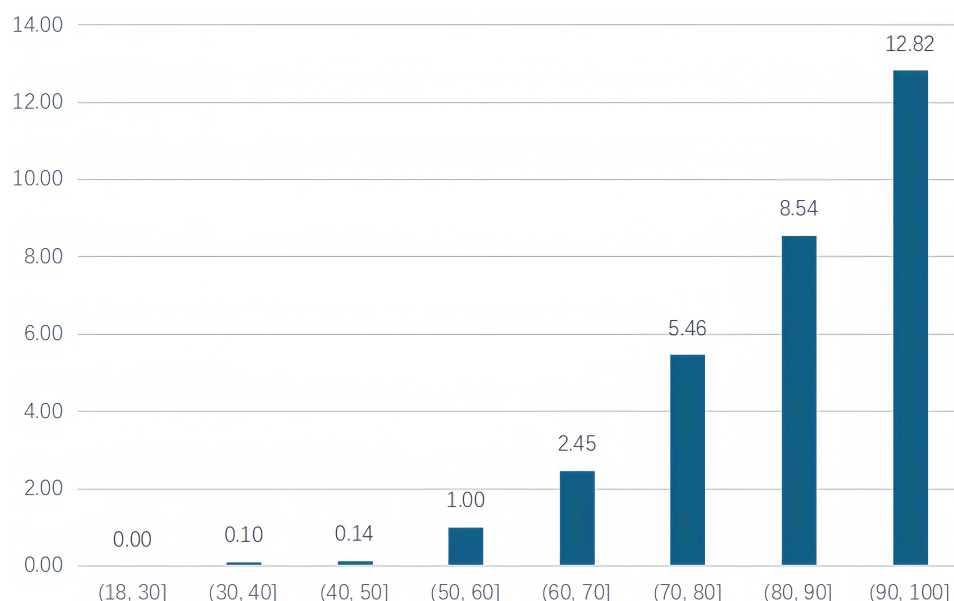

Figure 3 The incidence of lower extremity deep vein thrombosis by age groups. This figure depicted the incidence rate of deep vein thrombosis (DVT) in the lower limbs, expressed as a percentage, across various age groups. The data was categorized into age ranges, as indicated on the x-axis, and the corresponding incidence rates were plotted on the y-axis. Age Groups: The age groups were represented by intervals, such as (18,30], (30,40], (40,50], (50,60], (60,70], (70,80], (80,90], and (90,100], where the numbers in parentheses denoted the lower and upper bounds of each age range. The y-axis displayed the incidence of DVT, ranging from 0.00% to 14.00%.

Predictive Modelling and Model Performance Evaluation

Based on the 16 identified features, logistic regression, random forest, support vector machine, and XGBoost models were utilized for training and evaluation. A comparison of the overall performance of these models was presented in Table 3, with the ROC curves illustrated in Figure 4. Logistic regression and random forest exhibited superior

Table 2 LASSO Feature Screening

Characteristics	Coefficients
D-dimer	0.027896
Age	0.006531
Troponin T	0.004862
Urea Nitrogen	0.000782
Alzheimer's Disease	0.002066
Erythrocyte Count	−0.00054
Myoglobin	0.000772
Follicle Stimulating Hormone	0.000755
Porphyrin (Phosphorus)	−0.000585
Hemoglobin	−0.000348
Cholesterol	0.003517
Cholinesterase	−0.000506
Glucose	0.001775
Albumin	0.001686
C-reactive protein	−0.000149
Creatinine	−0.000818

Abbreviation: LASSO, Least Absolute Shrinkage and Selection Operator.

Table 3 Comprehensive Performance of Each Machine Learning Prediction Model

Model	Accuracy (95% CI)	Precision (95% CI)	Recall (95% CI)	F1 Score (95% CI)	AUC (95% CI)
Logistic Regression	0.851 (0.840, 0.862)	0.776 (0.765, 0.787)	0.918 (0.905, 0.931)	0.841 (0.828, 0.854)	0.900 (0.889, 0.911)
Random Forest	0.833 (0.820, 0.846)	0.750 (0.735, 0.765)	0.918 (0.905, 0.931)	0.826 (0.812, 0.840)	0.885 (0.873, 0.897)
Support Vector Machine	0.789 (0.776, 0.802)	0.736 (0.722, 0.750)	0.796 (0.783, 0.809)	0.765 (0.751, 0.779)	0.890 (0.878, 0.902)
ExtremeGradient Boosting	0.825 (0.812, 0.838)	0.730 (0.715, 0.745)	0.939 (0.926, 0.952)	0.821 (0.808, 0.834)	0.889 (0.877, 0.901)

performance, demonstrating high accuracy and recall rates, thus indicating their effectiveness in identifying the majority of positive class samples. The F1 scores and AUC further demonstrated the models' effectiveness in managing unbalanced data. The SVM model showed relatively weaker performance, particularly in terms of accuracy and F1 scores, indicating reduced effectiveness with unbalanced data. In contrast, the XGBoost model excelled in recall, proving effective at identifying most positive class samples. Although its precision was somewhat lower, the F1 score and AUC highlighted the model's strong performance in handling unbalanced datasets.

Feature Importance Analysis
In this study, a random forest model was employed to rank the relative importance of risk factors associated with the development of lower extremity deep vein thrombosis in hospitalized patients with mental disorders. The results indicated that D-dimer and age were the most influential variables, having the greatest impact on the predicted outcomes. The features were ranked in descending order of importance as follows: D-dimer (25.35%), age (20.18%), follicle-stimulating hormone (9.49%), troponin T (9.12%), and urea nitrogen (4.27%). The relative importance of each feature was illustrated in Figure 5.

Discussion
The objective of this study was to investigate the incidence of DVT and its influencing factors among psychiatric inpatients. To achieve this, we conducted a retrospective analysis of data from 17,434 psychiatric inpatients. The study found that the incidence of DVT in this population was 1.6%. Previous studies on various inpatient populations have reported DVT incidence rates ranging from 0.86% to 50%, and the incidence in our study population is higher than
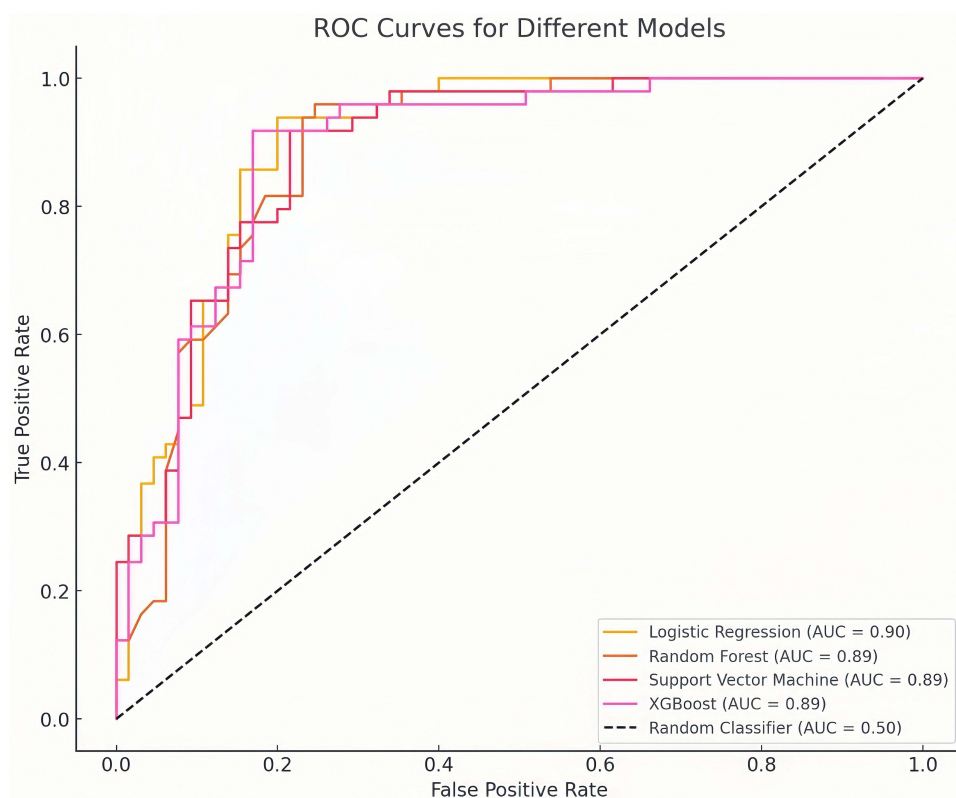
Figure 4 The ROC Curves Comparison for Various Machine Learning Models. This figure presented the Receiver Operating Characteristic (ROC) curves for four different machine learning models, along with a random classifier, to evaluate their performance in distinguishing between two classes. The ROC curves plotted the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The Area Under the Curve (AUC) provided a single measure of the model's ability to discriminate between the classes, with higher AUC values indicating better performance. Logistic Regression: AUC = 0.90. Random Forest: AUC = 0.89. Support Vector Machine: AUC = 0.89. XGBoost: AUC = 0.89. Random Classifier: AUC = 0.50. The x-axis represented the false positive rate (FPR), ranging from 0.0 to 1.0, and the y-axis represented the true positive rate (TPR), also ranging from 0.0 to 1.0. The diagonal line represented the performance of a random classifier, which serves as a baseline for comparison. Models with curves closer to the top-left corner of the plot demonstrated superior classification capabilities.

that observed in the general population and non-surgical inpatient populations.[7,8,15–17] These findings suggest that heightened vigilance and monitoring for thrombotic events are necessary for psychiatric inpatients, particularly those with multiple risk factors.

The machine learning models exhibited strong predictive performance overall, with each model displaying unique characteristics tailored to different clinical scenarios and needs. This offers a variety of options for personalized risk assessment. The study identified several key factors significantly associated with DVT, including D-dimer levels, age, Alzheimer's disease, and the use of Madopar. Furthermore, the significance of these factors was validated within the machine learning models.

Many conventional DVT scoring tools, such as the Oudega, Gagne, Wells, and Caprini scores, are limited by their lack of specificity in predictive power. These tools often rely on data, including specific blood indicators, that may not be readily accessible, which can undermine the accuracy of the predictive model. Consequently, the effectiveness of risk stratification in the inpatient setting was constrained by several factors.[18–22] Several retrospective studies have utilized machine learning techniques to predict DVT in various patient populations. Most of these studies have further segmented hospitalized patients into subgroups, such as those undergoing surgery, routine hospitalizations, oncology patients, and pregnant individuals. The predictive performance of machine learning models has consistently shown to be excellent across these diverse populations.[7,8,19,23,24]

This study highlights the effectiveness of machine learning in predicting the occurrence of DVT among psychiatric inpatients. Of the various techniques evaluated, the logistic regression model is favored for clinical risk assessment due to its clear interpretability and high accuracy. Additionally, the random forest model enhances predictive precision by
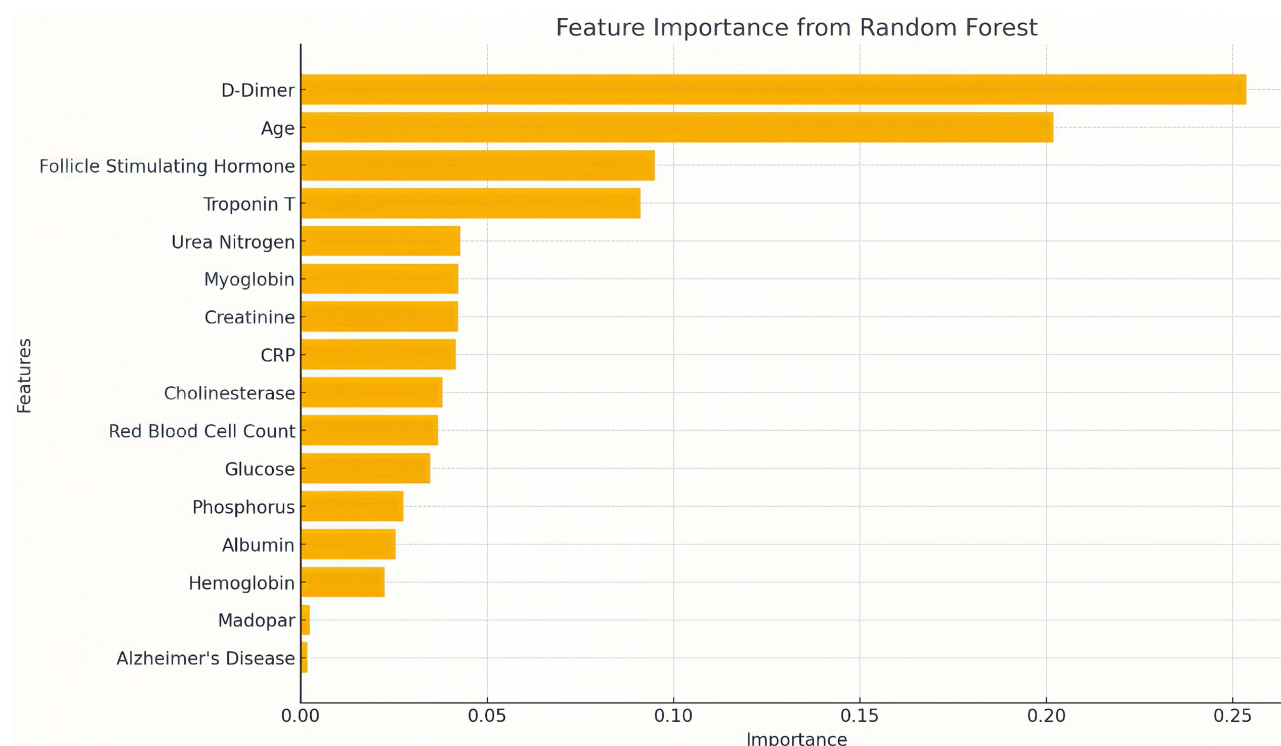
Feature Importance from Random Forest

Figure 5 The importance ranking of feature variables of random forest model. The illustration demonstrated the relative significance of the distinctive variables within the random forest model, as determined by their impact on the precision of the model predictions. The horizontal axis depicted the importance of the characteristic variables, with larger values indicating a greater impact on the model predictions. The vertical axis represented the characteristic variables, which include biomarkers and clinical parameters such as D-dimer, age, follicle-stimulating hormone, troponin T, urea nitrogen, myoglobin, creatinine, CRP, cholinesterase, erythrocyte count, glucose, phosphorus, albumin, hemoglobin, Madopar and Alzheimer's disease status.

leveraging its robust feature selection capabilities, which both confirms the importance of key influencing factors and improves prediction accuracy. The XGBoost model proved effective in managing data imbalance, particularly with respect to recall, underscoring its potential for identifying high-risk patients. In conclusion, machine learning models present a powerful tool for predicting DVT risk and supporting clinical decision-making in psychiatric inpatient populations. Their broad applicability is particularly valuable in clinical settings where integrating multiple influencing factors is essential for enhancing prediction accuracy.

This study identified several key clinical features, including D-dimer, age, follicle-stimulating hormone (FSH), and troponin T, which could be instrumental in developing predictive models for DVT. D-dimer, a fibrin degradation product commonly used as a biomarker for thrombosis, has been shown to be significantly associated with DVT development. Elevated D-dimer levels were a crucial predictor of DVT onset, as demonstrated by research across both general and specific patient populations.[15,17] The prominence of D-dimer in predictive models underscores its importance and suggests that it should be prioritized for inclusion in future clinical decision support systems.

Age is a non-modifiable risk factor that plays a significant role in the development of venous thromboembolism (VTE). Evidence suggests that the risk of developing DVT rises significantly with age, likely due to physiological changes such as decreased vascular elasticity, reduced blood flow, and lower levels of physical activity in older adults.[25] Elderly psychiatric inpatients are particularly vulnerable due to the presence of multiple chronic conditions that further impair mobility and elevate DVT risk. Given these considerations, it is advisable for clinical practice to emphasize the monitoring of geriatric psychiatric inpatients and to implement more proactive preventive measures to mitigate their risk of DVT.

Studies have identified a significant association between psychiatric disorders, including Alzheimer's disease, and the risk of developing venous thrombotic events, particularly among patients using Madopar. As Alzheimer's disease progresses, patients commonly experience cognitive decline and diminished physical mobility. The study's findings reinforced the substantial link between Alzheimer's disease and the development of DVT.[26] This highlights the

importance of addressing not only the cognitive aspects of Alzheimer's disease but also the overall physical health of patients, with particular attention to the risk of thrombosis, in their management and treatment.

The findings reveal an increased risk of DVT in patients using Madopar, a medication commonly prescribed for Parkinson's disease. This result aligned with previous research identifying knee/trunk flexion, a characteristic manifestation of Parkinson's disease (PD), and increased blood pressure variability as specific risk factors for DVT, along with lower limb movement limitations or hypokinesis.[27,28] Additionally, some studies suggested that antipsychotic medications might elevate the risk of DVT; however, these occurrences were relatively rare. The current study does not find substantial evidence supporting a significant impact of antipsychotic medications on the incidence of DVT events among psychiatric inpatients.[29–31]

These findings establish a basis for the early identification of DVT and guide the development of personalized prevention strategies. By integrating patient-specific information, clinicians can implement targeted preventive measures, thereby reducing the incidence of DVT and enhancing the overall prognosis for patients.

This study, a retrospective analysis based on a real-world database, has several limitations. Firstly, data were sourced from a single hospital, which, despite a large sample size, may limit the external validity due to the specific geographical location and patient population of that institution. Secondly, while the study included a diverse patient cohort, it did not account for potential confounding factors such as patients' lifestyles, nutritional status, and family history, which could influence the etiology of DVT. Future research should incorporate these variables to provide a more comprehensive understanding of DVT mechanisms and identify additional influencing factors. Thirdly, the low incidence of thrombotic events led to data balancing through undersampling, potentially resulting in the loss of valuable information and affecting the predictive accuracy of the model. Fourthly, since medical professionals determine the need for bilateral lower extremity venous ultrasound based on clinical judgment, this study might have underestimated the incidence of DVT. Fifth, the study excluded patients who died during their hospital stay, which may have led to an underestimation of the true incidence of DVT, as severe cases of DVT or PE could have contributed to in-hospital mortality. The exclusion of these patients might also have resulted in a selection bias. Finally, this study did not specifically examine the use of VTE prophylactic drugs, such as low-molecular-weight heparin (LMWH) or direct oral anticoagulants (DOACs), which are commonly used in clinical practice to prevent venous thromboembolism. The lack of data on these drugs may have impacted the results, as they could act as protective factors against DVT and PE.

Furthermore, the study was limited to lower-limb DVT, which likely underrepresents the true impact of venous thromboembolism (VTE) in psychiatric inpatients, as VTE can occur in other locations, such as the upper extremities or visceral veins, and pulmonary embolism (PE) remains a significant complication. Therefore, the actual burden of VTE in this population may be greater than what is reflected by this study's findings.

Despite these limitations, the study provides valuable insights into DVT incidence among psychiatric inpatients and influencing factors. Future studies should address these limitations to validate and expand upon these findings, including incorporating data on VTE prophylactic treatments, the effects of in-hospital mortality on DVT incidence, and the broader scope of VTE, including PE and thrombosis in other locations.

Conclusions

This study aimed to investigate the incidence of DVT and its key determinants among psychiatric inpatients using various machine learning models. The results showed that the incidence of lower extremity DVT in this population was 1.6%. D-dimer levels and age were identified as particularly significant predictors of DVT occurrence. Furthermore, the machine learning models demonstrated a high degree of accuracy in predicting DVT risk. The study also found that the incidence of psychiatric disorders and the use of medications such as Madopar in patients with Alzheimer's disease were associated with an increased risk of DVT. These findings highlight the importance of enhanced clinical vigilance in managing these patients. Overall, the study provides valuable insights for assessing the risk of venous thromboembolism (VTE) in psychiatric inpatients and offers a scientific basis for developing personalized prevention and treatment strategies.

Data Sharing Statement

Not applicable (this manuscript does not report data generation or analysis).

Ethics Approval and Consent to Participate

This retrospective analysis was conducted using data sourced from a real-world database, with ethical approval granted by the Ethics Committee of Huzhou Third Municipal Hospital (Ethics No. 2022SY2022084). Our study complies with the Declaration of Helsinki.

The relevant medical statistics data concerning hospitalized patients were anonymized, and, in accordance with the Ethics Committee of Huzhou Third Municipal Hospital approval, consent was obtained without the need for individual informed consent.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Collaborators GMD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9(2):137–150.
2. Navarrete S, Solar C, Tapia R, et al. Pathophysiology of deep vein thrombosis. *Clin Exp Med*. 2023;23(3):645–654. doi:10.1007/s10238-022-00829-w
3. Wang Z, Yang Y, He X, et al. Incidence and clinical features of venous thromboembolism in inpatients with Mental Illness. *Clin Appl Thromb/Hemost*. 2023;29:1299625039. doi:10.1177/10760296231160753
4. Kowal C, Peyre H, Amad A, et al. Psychotic, mood, and anxiety disorders and venous thromboembolism: a systematic review and meta-analysis. *Psychosomatic Med*. 2020;82(9):838–849. doi:10.1097/PSY.0000000000000863
5. Jørgensen H, Horváth-Puhó E, Laugesen K, et al. Venous thromboembolism and risk of depression: a population-based cohort study. *J Thromb Haemost*. 2023;21(4):953–962. doi:10.1016/j.jtha.2022.12.006
6. Duffett L. Deep venous thrombosis. *Ann Int Med*. 2022;175(9):ITC129–ITC144. doi:10.7326/AITC202209200
7. Wang X, Xi H, Geng X, et al. Artificial intelligence-based prediction of lower extremity deep vein thrombosis risk after knee/Hip arthroplasty. *Clin Appl Thromb/Hemost*. 2023;29:1309646529. doi:10.1177/10760296221139263
8. Shohat N, Ludwick L, Sherman MB, et al. Using machine learning to predict venous thromboembolism and major bleeding events following total joint arthroplasty. *Sci Rep*. 2023;13(1):2197. doi:10.1038/s41598-022-26032-1
9. Liu H, Yuan H, Wang Y, et al. Prediction of venous thromboembolism with machine learning techniques in young-middle-aged inpatients. *Sci Rep*. 2021;11(1):12868. doi:10.1038/s41598-021-92287-9
10. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci*. 2021;2(3):160. doi:10.1007/s42979-021-00592-x
11. Rajula HSR, Verlato G, Manchia M, et al. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*. 2020;56(9):455. doi:10.3390/medicina56090455
12. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517–518. doi:10.1001/jama.2017.7797
13. Zhang S, Chu W, Wang H, et al. Evaluation of stability of deep venous thrombosis of the lower extremities using Doppler ultrasound. *J Int Med Res*. 2020;48(8):1220741650.
14. Devi D, Biswas SK, Purkayastha B. A review on solution to class imbalance problem. In: Undersampling approaches: 2020 international conference on computational performance evaluation (ComPE), 2020. IEEE.
15. Jin S, Qin D, Liang B, et al. Machine learning predicts cancer-associated deep vein thrombosis using clinically available variables. *Int J Med Inf*. 2022;161:104733. doi:10.1016/j.ijmedinf.2022.104733
16. Lei H, Zhang M, Wu Z, et al. Development and validation of a risk prediction model for venous thromboembolism in lung cancer patients using machine learning. *Front Cardiovasc Med*. 2022;9:845210. doi:10.3389/fcvm.2022.845210

17. Meng L, Wei T, Fan R, et al. Development and validation of a machine learning model to predict venous thromboembolism among hospitalized cancer patients. *Asia-Pacific J Oncol Nurs*. 2022;9(12):100128. doi:10.1016/j.apjon.2022.100128

18. Wang P, Wang Y, Yuan Z, et al. Venous thromboembolism risk assessment of surgical patients in Southwest China using real-world data: establishment and evaluation of an improved venous thromboembolism risk model. *BMC Med Inf Decision Making*. 2022;22(1):59. doi:10.1186/s12911-022-01795-9

19. He L, Luo L, Hou X, et al. Predicting venous thromboembolism in hospitalized trauma patients: a combination of the Caprini score and data-driven machine learning model. *BMC Emerg Med*. 2021;21(1):60. doi:10.1186/s12873-021-00447-x

20. Ryan L, Mataraso S, Siefkas A, et al. A machine learning approach to predict deep venous thrombosis among hospitalized patients. *Clin Appl Thromb/Hemost*. 2021;27:1419577583. doi:10.1177/1076029621991185

21. Kafeza M, Shalhoub J, Salooja N, et al. A systematic review of clinical prediction scores for deep vein thrombosis. *Phlebology*. 2017;32(8):516–531. doi:10.1177/0268355516678729

22. Silveira PC, Ip IK, Goldhaber SZ, et al. Performance of Wells score for deep vein thrombosis in the inpatient setting. *JAMA Int Med*. 2015;175(7):1112–1117. doi:10.1001/jamainternmed.2015.1687

23. Wang KY, Ikwuezunma I, Puvanesarajah V, et al. Using predictive modeling and supervised machine learning to identify patients at risk for venous thromboembolism following posterior lumbar fusion. *Global Spine J*. 2023;13(4):1097–1103. doi:10.1177/21925682211019361

24. Nudel J, Bishara AM, de Geus SW, et al. Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. *Surg Endoscopy*. 2021;35:182–191. doi:10.1007/s00464-020-07378-x

25. Chen F, Xiong JX, Zhou WM. Differences in limb, age and sex of Chinese deep vein thrombosis patients. *Phlebology*. 2015;30(4):242–248. doi:10.1177/0268355514524192

26. Masters CL, Bateman R, Blennow K, et al. Alzheimer's disease. *Nat Rev Dis Primers*. 2015;1(1):1–18. doi:10.1038/nrdp.2015.56

27. Yamane K, Kimura F, Unoda K, et al. Postural abnormality as a risk marker for leg deep venous thrombosis in Parkinson's disease. *PLoS One*. 2013;8(7):e66984. doi:10.1371/journal.pone.0066984

28. Takeda T, Koreki A, Kokubun S, et al. Deep vein thrombosis and its risk factors in neurodegenerative diseases: a markedly higher incidence in Parkinson's disease. *J Neurol Sci*. 2024;457:122896. doi:10.1016/j.jns.2024.122896

29. Knudson JF, Kortepeter C, Dubitsky GM, et al. Antipsychotic drugs and venous thromboembolism. *Lancet*. 2000;356(9225):252–253. doi:10.1016/S0140-6736(05)74504-9

30. Waage IM, Gedde-Dahl A. Drug points: pulmonary embolism possibly associated with olanzapine treatment. *BMJ*. 2003;327(7428):1384. doi:10.1136/bmj.327.7428.1384

31. Liperoti R, Pedone C, Lapane KL, et al. Venous thromboembolism among elderly patients treated with atypical and conventional antipsychotic agents. *Arch Int Med*. 2005;165(22):2677–2682. doi:10.1001/archinte.165.22.2677