Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Systematic review of artificial intelligence for personalized oral anticoagulation management in cardiovascular care

Ahmed Alaqab[1]
University of Cyberjaya
Rosnani Binti Hashim[1]
rosnani@cyberjaya.edu.my

University of Cyberjaya
Maali Mujahid Mustafa[1]
University of Cyberjaya
Nurul Adilla Hayat Jamaluddin
University of Cyberjaya
Farizan Binti Abdul Ghaffar
Hospital Kuala Lumpur

Research Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11296432/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence and machine learning are increasingly used to personalize oral anticoagulation. We systematically reviewed patient-level studies in adults actively receiving warfarin, vitamin K antagonists, or direct oral anticoagulants. PubMed, Europe PMC, and Scopus were searched from inception to 15 August 2026. Of 984 records, 54 studies met eligibility criteria after duplicate removal, title and abstract screening, and full-text assessment. Applications included dose optimization and anticoagulation control, prediction of bleeding and thromboembolic outcomes, and individualized treatment selection. Warfarin dominated dose-optimization research, whereas direct oral anticoagulants were more prominent in risk and comparative-treatment modelling. Across the evidence base, limited external validation, inconsistent calibration, and sparse prospective evaluation restricted clinical translation. AI-enabled anticoagulation is progressing from static prediction toward longitudinal and treatment-specific decision support, but robust multicentre validation remains essential.

Introduction

Oral anticoagulants are central to the prevention and treatment of thromboembolic disease, particularly in atrial fibrillation (AF), venous thromboembolism, mechanical heart valves, and other high-risk thrombotic conditions [1–2]. Despite their established clinical benefit, anticoagulant therapy remains difficult to individualize because treatment response reflects a dynamic balance between thromboembolic protection and bleeding risk. Warfarin is especially challenging because of its narrow therapeutic index, interindividual variability in dose requirements, drug-drug and drug-food interactions, and the need for regular international normalized ratio (INR) monitoring. Direct oral anticoagulants (DOACs) simplify several aspects of management but do not eliminate clinically important heterogeneity in bleeding risk, thromboembolic protection, dose suitability, and treatment response.

Conventional anticoagulation management relies on clinician judgment, laboratory monitoring, population-derived dosing algorithms, and clinical risk scores. These approaches improve standardization but may not fully capture nonlinear interactions among age, renal and hepatic function, comorbidity, concurrent medication, genetic variation, and longitudinal treatment history. For warfarin, CYP2C9 and VKORC1 remain important determinants of dose requirements, while dynamic INR trajectories provide additional information that cannot be represented by a single baseline estimate [1, 6]. In DOAC-treated patients, risk scores such as HAS-BLED provide useful clinical structure but may have limited discrimination in specific treated populations [7, 24, 37].

Artificial intelligence and machine learning provide a computational framework for integrating high-dimensional clinical, laboratory, pharmacogenomic, and longitudinal data. Applications in anticoagulation now include random forests, gradient boosting, support-vector machines, artificial and deep neural networks, recurrent models, explainable AI, reinforcement learning, causal ML, and heterogeneous-treatment-effect methods [3–6]. These methods differ fundamentally in purpose: some predict a dose or outcome under a given therapy, whereas others attempt to recommend a sequential dose or estimate how treatment benefit and harm vary between alternative anticoagulants.

Warfarin dosing has been the most extensively investigated application. Recent evidence suggests that ML-based dose models can outperform conventional approaches within individual datasets, particularly when clinical and pharmacogenomic information is combined [6]. More recent longitudinal and reinforcement-learning studies extend beyond static dose estimation by learning from serial INR and dosing histories. At the same time, AI applications have expanded to bleeding and thromboembolic risk prediction during active anticoagulation and to individualized selection among warfarin and DOAC strategies [7–8, 24, 37, 39].

Despite rapid methodological development, important evidence gaps remain. Published studies vary markedly in population, clinical indication, model architecture, predictor availability, validation strategy, outcome definition, and performance reporting. High internal discrimination or low prediction error does not necessarily imply transportability, calibration, clinical utility, or improved patient outcomes. In addition, prediction studies and causal treatment-effect studies require different methodological assumptions and therefore cannot be judged using a single performance metric or risk-of-bias framework.

Accordingly, this systematic review aimed to evaluate patient-level AI and ML for personalized oral anticoagulation management in adults actively receiving oral anticoagulants. We specifically examined three clinical applications: (1) dose optimization and anticoagulation control, (2) prediction of clinically actionable bleeding and thromboembolic outcomes during treatment, and (3) individualized anticoagulant or dose selection. We further examined model characteristics, validation strategies, predictive performance, explainability, clinical utility, and methodological quality to identify the most important barriers to clinical translation.

Results

Study selection

The database search identified 984 records, including 227 from PubMed, 203 from Europe PMC, and 554 from Scopus. After removal of 405 duplicate records, 579 unique records remained for record-level screening. A total of 507 records were excluded before report retrieval (504 at title/abstract screening and 3 during final eligibility verification), leaving 72 reports sought for full-text retrieval. One report could not be retrieved, and 71 full-text reports were assessed for eligibility. Seventeen reports were excluded after full-text assessment because of an ineligible population (n = 6), ineligible AI/ML intervention (n = 5), ineligible outcome (n = 4), or ineligible study design/data provenance (n = 2). Ultimately, 54 studies met the eligibility criteria and were included in the systematic review. The study-selection process is summarized in Fig. 1, and the complete report-level screening ledger is provided in Supplementary Data 1, Table S2.

Characteristics of included studies

A total of 54 studies were included. The evidence base was dominated by warfarin/VKA applications, with additional studies evaluating DOAC-treated populations and mixed oral-anticoagulant cohorts. Clinical applications ranged from initial or maintenance warfarin dose prediction and longitudinal INR forecasting to prediction of major bleeding, intracranial hemorrhage, ischemic stroke, residual thrombotic risk, and individualized comparisons among anticoagulant strategies. Study designs were predominantly retrospective model-development or secondary-data analyses, although external validation cohorts, randomized-trial secondary analyses, and selected clinically oriented evaluations were also represented. Sample sizes ranged from small pharmacogenomic cohorts to large registry, claims, and randomized-trial datasets. Detailed characteristics are provided in Table 1.

The included studies used a broad spectrum of AI methodologies. Tree-based and ensemble methods included Random Forest, gradient boosting, XGBoost, LightGBM, CatBoost, stacking, and related ensembles. Kernel-based methods included support-vector regression and classification. Neural approaches included feedforward networks, deep neural networks, LSTM and bidirectional LSTM models, while sequential decision-making studies used reinforcement learning and offline policy learning. More recent treatment-selection studies incorporated causal ML, heterogeneous-treatment-effect learners, phenotype clustering, and digital-twin methods. Explainability approaches included SHAP, LIME, permutation importance, and conventional feature-importance measures.

Performance across clinical domains

Across the included studies, model performance varied according to clinical task, study population, predictor availability, outcome definition, and validation strategy (Table 2). Absolute performance metrics should therefore be interpreted within the prediction target of each study rather than compared directly across heterogeneous tasks. Dose-prediction studies reported metrics such as MAE, RMSE, R^2, PW20, or dose-classification accuracy, whereas clinical-event studies more commonly reported AUC/C-index, sensitivity, specificity, calibration, or risk reclassification. Treatment-effect studies reported hazard ratios, risk differences, interaction tests, or individualized treatment-effect estimates.

Table 2
Reported performance outcomes

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
Warfarin Dose Estimation on High-dimensional and Incomplete Data (2021) [16]	Warfarin/VKA	Model development/evaluation study N = 5,410 included patients	Feature embedding + multi-instance neural network with attention/multi-head pooling	Baseline ML models	Warfa dose-p perfor
Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method (2023) [17]	Apixaban; dabigatran; rivaroxaban; warfarin	Retrospective claims cohort; causal ML N = 34,569 patients	Causal machine learning after ML matching	Head-to-head OAC comparisons	Comp stroke hemor morta
Evaluating warfarin dosing models on multiple datasets with a novel software framework and evolutionary optimisation (2021) [18]	Warfarin/VKA	Algorithm evaluation across two datasets N=IWPC 5,741; PathCare 4,621	Multiple regression/ML algorithms + evolutionary optimisation; Warfit-learn framework	Cross-model comparison	Weekl
Principled estimation and evaluation of treatment effect heterogeneity: A case study application to dabigatran for patients with atrial fibrillation (2023) [19]	Dabigatran 110 mg; dabigatran 150 mg; warfarin	Post hoc analysis of randomized trial data N = 17,282 participants	S-learner, X-learner, R-learner, causal survival forest	Three anticoagulant strategies	Stroke embol bleedi
A Prediction Model of Stable Warfarin Doses in Patients After Mechanical Heart Valve Replacement Based on a Machine Learning Algorithm (2025) [20]	Warfarin/VKA	Retrospective single-center model development/validation N = 413 patients	10 ML algorithms; best SVM radial-basis-function model	Traditional linear model; GLM; other ML models	Stable
A Comparative Analysis of Machine Learning Techniques for Therapeutic Warfarin Dose Prediction (2024) [21]	Warfarin/VKA	Preliminary ML model-development study N = 5,052 records	MLP, SVM, KNN, XGBoost, CatBoost, gradient boosting tree, RF, decision tree, logistic regression	Nine ML algorithms	Therap dose c
AI-assisted warfarin dose optimisation with CURATE.AI for clinical impact: Retrospective data analysis (2025) [22]	Warfarin/VKA	Retrospective data analysis N = 127 patients	CURATE.AI personalized dose-INR response profiling	Physician-guided dosing; literature models	Dose r INR re TTR
Stable warfarin dose prediction in sub-Saharan African patients: A machine-learning approach and external validation of a clinical dose-initiation algorithm (2022) [23]	Warfarin/VKA	Multicenter observational model development + external validation N = 364 development; 270 external validation	21 ML techniques; random forest regression among top approaches	Clinical dose-initiation algorithm; multiple ML techniques	Stable dose
Residual risk prediction in anticoagulated patients with atrial fibrillation using machine learning: A report from the GLORIA-AF registry phase II/III (2025) [24]	Oral anticoagulants (mixed)	Registry secondary analysis N = 15,829 patients	Logistic regression, random forest, LightGBM, XGBoost	CHA2DS2-VA; 2MACE	Ischer syster MI cor
Machine Learning Predicts Bleeding Risk in Atrial Fibrillation Patients on Direct Oral Anticoagulant (2025) [25]	DOACs	Retrospective EHR prognostic cohort N = 24,468 patients	Random forest, XGBoost and other ML models	HAS-BLED, ATRIA, ORBIT and other conventional scores	Hospi bleedi stroke
Heterogeneity of Treatment Effect of DOACs vs Warfarin in Atrial Fibrillation (2026) [26]	Warfarin; standard-dose DOACs	Pooled individual-participant analysis of 4 pivotal RCTs N = 58,634 patients	Clustering / gradient-boosting treatment-effect modeling	DOACs vs warfarin	Net cli includ bleedi

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
Evaluation of Machine Learning Approaches for Predicting Warfarin Discharge Dose in Cardiac Surgery Patients: Retrospective Algorithm Development and Validation Study (2023) [27]	Warfarin/VKA	Retrospective algorithm development/validation N = 1,031 encounters	Penalized linear regression, KNN, RF, gradient boosting, MARS, ensemble	Multiple ML approaches	Warfa for the discha
Risk prediction score for clinical outcome in atrial fibrillation and stable coronary artery disease (2023) [28]	Rivaroxaban-based AFIRE therapy	Post hoc randomized-trial analysis N = 2,215 patients	Random survival forest; Cox regression; Boruta feature selection	Conventional Cox-based score / clinical risk variables	NACE MI, str bleedi
Improved stacking ensemble learning based on feature selection to accurately predict warfarin dose (2024) [29]	Warfarin/VKA	Retrospective model-development study N = 641 patients	Heuristic stacking ensemble with feature-selection models	Traditional stacking / component models	Warfa
Intracardiac thrombus formations despite continuous oral anticoagulation in atrial fibrillation patients undergoing catheter ablation procedures: pilot development of a machine learning prediction model (2026) [30]	DOACs 96%; VKA 4%	Retrospective matched cohort; pilot ML development N = 149 modeled patients (65 TF+; 84 TF−); 1,335 screened	ML models using clinical, echocardiographic and P-wave features	CHA2DS2-VASc	Intraca throm forma
Machine Learning Algorithm for Predicting Warfarin Dose in Caribbean Hispanics Using Pharmacogenetic Data (2020) [31]	Warfarin/VKA	Secondary analysis of pharmacogenetic cohort N = 190 patients	Random forest regression, SVR, MARS and other ML models	Multiple ML models	Stable
Machine Learning for Prediction of Stable Warfarin Dose in US Latinos and Latin Americans (2021) [32]	Warfarin/VKA	Multicohort pharmacogenetic model study N = 7,030 total; ULLA 1,734	BART, MARS, SVR, linear models	IWPC algorithm / clinical model	Stable
Warfarin anticoagulation management during the COVID-19 pandemic: The role of internet clinic and machine learning (2022) [33]	Warfarin/VKA	Retrospective cohort N = 241 patients	XGBoost	Hospital anticoagulation clinic vs Internet clinic; conventional factors	TTR a anticoa
AI-WAR: a novel warfarin management software with a bidirectional LSTM dosing model improves time in therapeutic range (2026) [34]	Warfarin/VKA	Model training from RCT + prospective registry; external real-world validation N = 624 RCT + 176 registry for training; 200 external validation	LSTM and bidirectional LSTM within AI-WAR	Conventional management; unidirectional LSTM	TTR, ti adver predic
Integrated Machine Learning Decision Tree Model for Risk Evaluation in Patients with Non-Valvular Atrial Fibrillation When Taking Different Doses of Dabigatran (2023) [35]	Dabigatran 110 mg or 150 mg	Post hoc RE-LY analysis N = 11,802 dabigatran users (5,869 on 110 mg; 5,933 on 150 mg)	Naive Bayes, RF, CART, XGBoost; logistic regression benchmark	Logistic regression	Vascu mg); b
Important Risk Factors in Patients with Nonvalvular Atrial Fibrillation Taking Dabigatran Using Integrated Machine Learning Scheme—A Post Hoc Analysis (2022) [36]	Dabigatran 110 mg or 150 mg	Post hoc RE-LY analysis N = 11,802 patients with complete dabigatran follow-up	Naive Bayes, RF, CART, XGBoost; logistic regression	Logistic regression	Vascu bleedi

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
Development and Validation of a Predictive Model for Intracranial Haemorrhage in Patients on Direct Oral Anticoagulants (2024) [37]	DOACs	Retrospective electronic-health-data model development/validation N = 24,794 patients	XGBoost-based predictive model	Existing DOAC bleeding/risk scores	Intracı haemo
Machine learning guided prediction of warfarin blood levels for personalized medicine based on clinical longitudinal data from cardiac surgery patients: a prospective observational study (2024) [38]	Warfarin/VKA	Prospective observational study N = 246 patients; 3,261 INR measurements	RF plus multiple regression/classification ML algorithms	Multiple ML algorithms	INR / t classi contin
Explainable Artificial Intelligence Unravels the Possible Distinct Roles of VKORC1 and CYP2C9 in Predicting Warfarin Anticoagulation Control (2026) [39]	Warfarin/VKA	Post hoc cross-sectional dataset analysis N = 232 patients	Random Forest, XGBoost, logistic regression	Logistic regression	Poor A TTR <
Can Machine Learning from Real-World Data Support Drug Treatment Decisions? A Prediction Modeling Case for Direct Oral Anticoagulants (2022) [40]	Apixaban; rivaroxaban	Retrospective claims treatment-effect modeling N = 29,901 new users	Random forests with X-learner framework	Observed treatment / trial-emulation comparisons	Stroke compa
Gene Polymorphisms of the Renin-Angiotensin System and Bleeding Complications of Warfarin: Genetic-Based Machine Learning Models (2021) [41]	Warfarin/VKA	Observational pharmacogenetic study with predictive modeling N = 142 patients	Logistic regression, elastic net, RF, SVM linear/radial	Multiple model comparison	Warfa bleedi
Predictors of bleeding event among elderly patients with mechanical valve replacement using random forest model: A retrospective study (2021) [42]	Warfarin/VKA	Retrospective cohort N = 598 participants	Random forest	Clinical predictor analysis	Bleedi
Evolutionary synthetic oversampling technique and cocktail ensemble model for warfarin dose prediction with imbalanced data (2021) [43]	Warfarin/VKA	Model-development study using combined datasets N = 733 samples	ESMOTE + cocktail ensemble model using RF/BRT components	Other oversampling methods; RF, BRT, AdaBoost	Stable warfa
An Adapted Neural-Fuzzy Inference System Model Using Preprocessed Balance Data to Improve the Predictive Accuracy of Warfarin Maintenance Dosing in Patients After Heart Valve Replacement (2022) [44]	Warfarin/VKA	Retrospective multicenter model development/validation N = 15,108 patients	ANFIS on imbalanced vs balanced training data	Imbalanced vs balanced ANFIS	Warfa dose
Development of an interpretable machine learning model to predict short-term bleeding risk in patients receiving dual antithrombotic therapy	Warfarin + aspirin	Retrospective cohort with temporal external validation N = 499 development/internal; 93 external validation	LR, decision tree, RF, SVM, XGBoost, LightGBM, KNN	Seven ML models	Bleedi month

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
following cardiac surgery (2026) [45]					
XGBoost-based machine learning test improves the accuracy of hemorrhage prediction among geriatric patients with long-term administration of rivaroxaban (2023) [46]	Rivaroxaban	Longitudinal clinical follow-up cohort N = 798 patients	Logistic regression, RF, XGBoost	Logistic regression; RF	Bleedi
DBCSMOTE: a clustering-based oversampling technique for data-imbalanced warfarin dose prediction (2020) [47]	Warfarin/VKA	Methodology/model-development study N = 592 samples (357 + 235)	DBCSMOTE + BRT / RF	Other oversampling approaches	Warfa
A Real-World Exploration into Clinical Outcomes of Direct Oral Anticoagulant Dosing Regimens in Morbidly Obese Patients Using Data-Driven Approaches (2023) [48]	Apixaban; rivaroxaban; dabigatran/other DOAC regimens	Retrospective real-world EHR study N = 4,275 patients	Decision tree, RF, bootstrap aggregation plus multivariable regression	Different DOAC dose regimens	Morta bleedi outcor
The Prediction Model of Warfarin Individual Maintenance Dose for Patients Undergoing Heart Valve Replacement, Based on the Back Propagation Neural Network (2020) [49]	Warfarin/VKA	Retrospective multicenter model development/validation N = 13,639 patients	Back Propagation Neural Network	Multiple linear regression	Warfa dose
Warfarin maintenance dose prediction for Chinese after heart valve replacement by a feedforward neural network with equal stratified sampling (2021) [50]	Warfarin/VKA	Retrospective multicenter model development/validation N = 19,060 eligible cases	Feedforward neural network with equal stratified sampling	Model trained on raw imbalanced dataset	Warfa dose
Development of a system to support warfarin dose decisions using deep neural networks (2021) [51]	Warfarin/VKA	Retrospective multi-institutional model development N = 19,719 inpatients; 22,314 training sequences and 12,673 external test sequences reported	Dense and recurrent neural networks	10 expert physicians	5th-da predic dose-l
Machine learning models to predict the warfarin discharge dosage using clinical information of inpatients from South Korea (2023) [52]	Warfarin/VKA	Retrospective model development + external validation N = 3,168 development; 891 external validation	Four ML models including XGBoost/ANN/RF (paper evaluates multiple algorithms)	Physician dose predictions	Warfa
Optimizing warfarin dosing for patients with atrial fibrillation using machine learning (2024) [53]	Warfarin/VKA	Secondary analysis of pivotal RCT warfarin arms; deep RL policy development N = 22,502 development; 5,730 external validation	Batch-Constrained deep Q-learning / deep reinforcement learning	Rules-based clinical dosing benchmark; observed clinician dosing	TTR; c outcor
Retrospective modeling study of ChatGPT (GPT-4) warfarin dose adjustment in patients with INRs outside the therapeutic range (2026) [54]	Warfarin/VKA	Retrospective modeled study N = 180 patients	ChatGPT (GPT-4)	Clinician recommendations	Weekl INR no
Long Short-Term Memory Network for Development and Simulation of Warfarin Dosing Model Based on Time Series	Warfarin	Retrospective time-series model development and external validation N = 624 development; 158 external validation	LSTM recurrent neural network	Clinical/MAPB dosing algorithm and observed dosing	Longit predic simula

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
Anticoagulant Data (2022) [55]					
Automated warfarin dose prediction for Asian, American, and Caucasian populations using a deep neural network (2023) [56]	Warfarin	Secondary patient-level model development/validation N = > 6,000 IWPC records	Deep neural network	Published warfarin dosing algorithms / population-specific models	Indivic warfa
Using machine learning to predict the bleeding risk for patients with cardiac valve replacement treated with warfarin in hospitalized (2024) [57]	Warfarin	Retrospective hospitalized cohort N = 2,376	Five ML models; XGBoost best-performing model	Alternative ML models / conventional assessment	High b during (INR ≥
Predictors of warfarin therapeutic dose across thromboembolic indications: an exploratory early clinical evaluation study using machine learning-integrated explainable artificial intelligence and Bayesian models (2026) [58]	Warfarin	Exploratory patient-level ML/XAI/Bayesian analysis N=Clinical/pharmacogenomic patient-level dataset	Five ML models + SHAP + BART	Alternative ML and Bayesian models	Therap dose a and D
A Post-Hoc Interpretable Ensemble Model to Feature Effect Analysis in Warfarin Dose Prediction for Chinese Patients (2022) [59]	Warfarin	Post-hoc model interpretation study N=Chinese warfarin cohort	Interpretable ensemble + permutation importance + LIME + SHAP	Component/baseline predictive models	Daily v
Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning (2022) [60]	Warfarin	Retrospective EHR reinforcement-learning study N = 10,408 total; test set n = 1,260	Reinforcement learning	Clinicians and alternative algorithms	Seque optimi respor
Warfarin Dose Management Using Offline Deep Reinforcement Learning (2025) [61]	Warfarin	Offline deep reinforcement-learning study N=Historical longitudinal warfarin cohort	Batch-Constrained Q-learning (BCQ)	XGBoost regression and LSTM baseline	Cumu next IN predic
Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation (2024) [62]	VKA and DOAC	Anticoagulated AF prognostic ML cohort N = 11,078	Multi-Gate Mixture of Experts, Gradient Boosting and other ML models	CHA_2DS_2-VASc and HAS-BLED	Major ischer cause
An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set (2021) [63]	Warfarin	Observational model-development study N = 377 patients; 1,173 warfarin order events	LightGBM ensemble framework with cross-over variables	Baseline models and IWPC model	Warfa dose
Optimizing warfarin dosing in diabetic patients through BERT model and machine learning techniques (2025) [64]	Warfarin	Secondary patient-level ML model-development study N=IWPC cohort; diabetic subgroup analysed	BERT-informed representation + RF, KNN, MLP, SVM, linear regression	Linear regression and alternative ML models	Indivic dose i diabet
A Digital Twin Strategy to Predict Thrombotic Recurrence in Antiphospholipid Syndrome Patients Treated with Direct Oral Anticoagulants vs. Vitamin K Antagonists	VKA vs DOAC	Multicenter observational digital-twin/causal modeling study N = 89 APS patients (70 VKA; 19 DOAC)	Conditional GAN digital twin + causal DAG	DOAC vs VKA	Throm and in counte treatm

Study	OAC	Design and sample	AI/ML model	Comparator	Outco
Using Data from Real-World Populations (2025) [65]					
Evaluation of supervised machine learning algorithms in predicting the poor anticoagulation control and stable weekly doses of warfarin (2023) [66]	Warfarin	Patient-level supervised ML prediction study N=Warfarin users treated ≥ 6 months	C5.0, CART and CHAID	Alternative tree-based algorithms	Poor a contro weekl
Determining the adjusted initial treatment dose of warfarin anticoagulant medicine using kernel-based support vector regression (2022) [67]	Warfarin	Patient-level computational dose-prediction study N=IWPC pharmacogenomic dataset	Kernel-based support vector regression	Conventional/predefined kernels	Adjust dose
Unsupervised clustering approach to assess heterogeneity of treatment effects across patient phenotypes in randomized clinical trials (2025) [68]	Higher-dose edoxaban vs warfarin	Secondary randomized-trial HTE analysis N = 14,062 ENGAGE AF-TIMI 48 participants	Unsupervised phenotype clustering	Edoxaban vs warfarin	Stroke embol effect
Predicting Range of Initial Warfarin Dose Based on Pharmacometabolomic and Genetic Inputs (2021) [69]	Warfarin	Pharmacometabolomic/pharmacogenetic patient-level model-development study N = 160 heart-valve patients	PLS + LDA/multinomial-logit ML + IniWarD	Observed INR response / dose-response strata	Day-7 initial

Study	Domain	Model	Outcome	Comparator	Validation
Warfarin Dose Estimation on High-dimensional and Incomplete Data (2021) [16]	Warfarin dose optimization	Feature embedding + multi-instance neural network with attention/multi-head pooling	Warfarin dose; subgroup dose-prediction performance	Baseline ML models	Not clearly external; benchmark/test evaluation
Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method (2023) [17]	Individualized OAC selection	Causal machine learning after ML matching	Composite ischemic stroke, intracranial hemorrhage, all-cause mortality	Head-to-head OAC comparisons	Internal observation evaluation
Evaluating warfarin dosing models on multiple datasets with a novel software framework and evolutionary optimisation (2021) [18]	Warfarin dose optimization	Multiple regression/ML algorithms + evolutionary optimisation; Warfit-learn framework	Weekly warfarin dose	Cross-model comparison	Repeated train/test evaluation across tv datasets
Principled estimation and evaluation of treatment effect heterogeneity: A case study application to dabigatran for patients with atrial fibrillation (2023) [19]	Individualized OAC selection	S-learner, X-learner, R-learner, causal survival forest	Stroke/systemic embolism; major bleeding	Three anticoagulant strategies	70/30 train-test evaluation with HTE metrics
A Prediction Model of Stable Warfarin Doses in Patients After Mechanical Heart Valve Replacement Based on a Machine Learning Algorithm (2025) [20]	Warfarin dose optimization	10 ML algorithms; best SVM radial-basis-function model	Stable warfarin dose	Traditional linear model; GLM; other ML models	Random 80/20 train validation split
A Comparative Analysis of Machine Learning Techniques for Therapeutic Warfarin Dose Prediction (2024) [21]	Warfarin dose optimization	MLP, SVM, KNN, XGBoost, CatBoost, gradient boosting tree, RF, decision tree, logistic regression	Therapeutic warfarin dose class	Nine ML algorithms	Train/test split; tuning/balancing described
AI-assisted warfarin dose optimisation with CURATE.AI for clinical impact: Retrospective data analysis (2025) [22]	Warfarin dose optimization	CURATE.AI personalized dose-INR response profiling	Dose recommendation; INR response; modeled TTR	Physician-guided dosing; literature models	Retrospective mode evaluation
Stable warfarin dose prediction in sub-Saharan African patients: A machine-learning approach and external validation of a clinical dose-initiation algorithm (2022) [23]	Warfarin dose optimization	21 ML techniques; random forest regression among top approaches	Stable weekly warfarin dose	Clinical dose-initiation algorithm; multiple ML techniques	External temporal/geographi validation; bootstrap tuning
Residual risk prediction in anticoagulated patients with atrial fibrillation using machine learning: A report from the GLORIA-AF registry phase II/III (2025) [24]	Residual thrombotic risk under OAC	Logistic regression, random forest, LightGBM, XGBoost	Ischemic stroke, systemic embolism, TIA, MI composite	CHA2DS2-VA; 2MACE	Train/test evaluatio
Machine Learning Predicts Bleeding Risk in Atrial Fibrillation Patients on Direct Oral Anticoagulant (2025) [25]	DOAC bleeding-risk prediction	Random forest, XGBoost and other ML models	Hospitalization for bleeding; hemorrhagic stroke	HAS-BLED, ATRIA, ORBIT and other conventional scores	Internal resampling/cross-validation; no independent externa cohort
Heterogeneity of Treatment Effect of DOACs vs Warfarin in Atrial Fibrillation (2026) [26]	Individualized OAC selection	Clustering / gradient-boosting treatment-effect modeling	Net clinical outcome including ischemic and bleeding events	DOACs vs warfarin	Internal modeling o pooled RCT data

Study	Domain	Model	Outcome	Comparator	Validation
Evaluation of Machine Learning Approaches for Predicting Warfarin Discharge Dose in Cardiac Surgery Patients: Retrospective Algorithm Development and Validation Study (2023) [27]	Warfarin dose optimization	Penalized linear regression, KNN, RF, gradient boosting, MARS, ensemble	Warfarin dose needed for therapeutic INR at discharge	Multiple ML approaches	10-fold CV and held- test cohort
Risk prediction score for clinical outcome in atrial fibrillation and stable coronary artery disease (2023) [28]	Clinical risk under OAC	Random survival forest; Cox regression; Boruta feature selection	NACE composite (death, MI, stroke, major bleeding)	Conventional Cox-based score / clinical risk variables	Development/valida split
Improved stacking ensemble learning based on feature selection to accurately predict warfarin dose (2024) [29]	Warfarin dose optimization	Heuristic stacking ensemble with feature-selection models	Warfarin dose	Traditional stacking / component models	90/10 train-test spli CV/tuning
Intracardiac thrombus formations despite continuous oral anticoagulation in atrial fibrillation patients undergoing catheter ablation procedures: pilot development of a machine learning prediction model (2026) [30]	Residual thrombotic risk under OAC	ML models using clinical, echocardiographic and P-wave features	Intracardiac thrombus/prethrombotic formation	CHA2DS2-VASc	External validation reported
Machine Learning Algorithm for Predicting Warfarin Dose in Caribbean Hispanics Using Pharmacogenetic Data (2020) [31]	Warfarin dose optimization	Random forest regression, SVR, MARS and other ML models	Stable warfarin dose	Multiple ML models	80/20 train-test spli
Machine Learning for Prediction of Stable Warfarin Dose in US Latinos and Latin Americans (2021) [32]	Warfarin dose optimization	BART, MARS, SVR, linear models	Stable warfarin dose	IWPC algorithm / clinical model	30% testing with 10 resampling rounds
Warfarin anticoagulation management during the COVID-19 pandemic: The role of internet clinic and machine learning (2022) [33]	Warfarin anticoagulation control	XGBoost	TTR and INR variability / anticoagulation quality	Hospital anticoagulation clinic vs Internet clinic; conventional factors	Internal model evaluation
AI-WAR: a novel warfarin management software with a bidirectional LSTM dosing model improves time in therapeutic range (2026) [34]	Warfarin dose/TTR optimization	LSTM and bidirectional LSTM within AI-WAR	TTR, time in target INR, adverse events, dose-prediction accuracy	Conventional management; unidirectional LSTM	Independent externa real-world cohort
Integrated Machine Learning Decision Tree Model for Risk Evaluation in Patients with Non-Valvular Atrial Fibrillation When Taking Different Doses of Dabigatran (2023) [35]	DOAC dose-specific risk	Naive Bayes, RF, CART, XGBoost; logistic regression benchmark	Vascular events (110 mg); bleeding (150 mg)	Logistic regression	90/10 split with cros validation/feature-selection workflow
Important Risk Factors in Patients with Nonvalvular Atrial Fibrillation Taking Dabigatran Using Integrated Machine Learning Scheme—A	DOAC bleeding/thrombotic risk	Naive Bayes, RF, CART, XGBoost; logistic regression	Vascular events; bleeding	Logistic regression	90/10 train-test spli

Study	Domain	Model	Outcome	Comparator	Validation
Post Hoc Analysis (2022) [36]					
Development and Validation of a Predictive Model for Intracranial Haemorrhage in Patients on Direct Oral Anticoagulants (2024) [37]	DOAC ICH risk	XGBoost-based predictive model	Intracranial haemorrhage	Existing DOAC bleeding/risk scores	Random 4:1 development-valida split
Machine learning guided prediction of warfarin blood levels for personalized medicine based on clinical longitudinal data from cardiac surgery patients: a prospective observational study (2024) [38]	Warfarin INR/control prediction	RF plus multiple regression/classification ML algorithms	INR / target-INR classification and continuous prediction	Multiple ML algorithms	Training/validation evaluation
Explainable Artificial Intelligence Unravels the Possible Distinct Roles of VKORC1 and CYP2C9 in Predicting Warfarin Anticoagulation Control (2026) [39]	Warfarin TTR/control prediction	Random Forest, XGBoost, logistic regression	Poor ACS defined as TTR < 70%	Logistic regression	70/30 train-test spli
Can Machine Learning from Real-World Data Support Drug Treatment Decisions? A Prediction Modeling Case for Direct Oral Anticoagulants (2022) [40]	Individualized DOAC selection	Random forests with X-learner framework	Stroke; major bleeding; composite	Observed treatment / trial-emulation comparisons	Trial-emulation-base internal evaluation
Gene Polymorphisms of the Renin-Angiotensin System and Bleeding Complications of Warfarin: Genetic-Based Machine Learning Models (2021) [41]	Warfarin bleeding-risk prediction	Logistic regression, elastic net, RF, SVM linear/radial	Warfarin-related bleeding complications	Multiple model comparison	Repeated 5-fold cro validation
Predictors of bleeding event among elderly patients with mechanical valve replacement using random forest model: A retrospective study (2021) [42]	Warfarin bleeding-risk prediction	Random forest	Bleeding events	Clinical predictor analysis	Internal model evaluation
Evolutionary synthetic oversampling technique and cocktail ensemble model for warfarin dose prediction with imbalanced data (2021) [43]	Warfarin dose optimization	ESMOTE + cocktail ensemble model using RF/BRT components	Stable therapeutic warfarin dose	Other oversampling methods; RF, BRT, AdaBoost	Train/test model evaluation
An Adapted Neural-Fuzzy Inference System Model Using Preprocessed Balance Data to Improve the Predictive Accuracy of Warfarin Maintenance Dosing in Patients After Heart Valve Replacement (2022) [44]	Warfarin dose optimization	ANFIS on imbalanced vs balanced training data	Warfarin maintenance dose	Imbalanced vs balanced ANFIS	Training + internal a external validation

Study	Domain	Model	Outcome	Comparator	Validation
Development of an interpretable machine learning model to predict short-term bleeding risk in patients receiving dual antithrombotic therapy following cardiac surgery (2026) [45]	Warfarin bleeding-risk prediction	LR, decision tree, RF, SVM, XGBoost, LightGBM, KNN	Bleeding within 3 months after discharge	Seven ML models	70/30 internal split + temporally independ external cohort
XGBoost-based machine learning test improves the accuracy of hemorrhage prediction among geriatric patients with long-term administration of rivaroxaban (2023) [46]	DOAC bleeding-risk prediction	Logistic regression, RF, XGBoost	Bleeding adverse events	Logistic regression; RF	Model comparison; additional validatior cohort reported in paper
DBCSMOTE: a clustering-based oversampling technique for data-imbalanced warfarin dose prediction (2020) [47]	Warfarin dose optimization	DBCSMOTE + BRT / RF	Warfarin daily dose	Other oversampling approaches	Training/validation/ split
A Real-World Exploration into Clinical Outcomes of Direct Oral Anticoagulant Dosing Regimens in Morbidly Obese Patients Using Data-Driven Approaches (2023) [48]	DOAC dose/outcome risk	Decision tree, RF, bootstrap aggregation plus multivariable regression	Mortality, stroke, bleeding and clinical outcomes	Different DOAC dose regimens	70/30 stratified trair test split
The Prediction Model of Warfarin Individual Maintenance Dose for Patients Undergoing Heart Valve Replacement, Based on the Back Propagation Neural Network (2020) [49]	Warfarin dose optimization	Back Propagation Neural Network	Warfarin maintenance dose	Multiple linear regression	Training + internal validation + 10% external validation
Warfarin maintenance dose prediction for Chinese after heart valve replacement by a feedforward neural network with equal stratified sampling (2021) [50]	Warfarin dose optimization	Feedforward neural network with equal stratified sampling	Warfarin maintenance dose	Model trained on raw imbalanced dataset	Internal + holdout external validation
Development of a system to support warfarin dose decisions using deep neural networks (2021) [51]	Warfarin INR/dose optimization	Dense and recurrent neural networks	5th-day PT-INR prediction; individualized dose-INR simulation	10 expert physicians	Cross-institution testing
Machine learning models to predict the warfarin discharge dosage using clinical information of inpatients from South Korea (2023) [52]	Warfarin dose optimization	Four ML models including XGBoost/ANN/RF (paper evaluates multiple algorithms)	Warfarin discharge dose	Physician dose predictions	Internal validation + external MIMIC-III validation
Optimizing warfarin dosing for patients with atrial fibrillation using machine learning (2024) [53]	Dynamic warfarin dosing	Batch-Constrained deep Q-learning / deep reinforcement learning	TTR; composite clinical outcomes	Rules-based clinical dosing benchmark; observed clinician dosing	External validation i RE-LY across 44 countries
Retrospective modeling study of ChatGPT (GPT-4) warfarin dose	Warfarin dose adjustment / LLM	ChatGPT (GPT-4)	Weekly dose; modeled INR normalization	Clinician recommendations	Retrospective concordance/mode only

Study	Domain	Model	Outcome	Comparator	Validation
adjustment in patients with INRs outside the therapeutic range (2026) [54]					
Long Short-Term Memory Network for Development and Simulation of Warfarin Dosing Model Based on Time Series Anticoagulant Data (2022) [55]	Dynamic warfarin dose/INR optimization	LSTM recurrent neural network	Longitudinal INR prediction and dose simulation	Clinical/MAPB dosing algorithm and observed dosing	External validation cohort
Automated warfarin dose prediction for Asian, American, and Caucasian populations using a deep neural network (2023) [56]	Warfarin dose optimization	Deep neural network	Individualized weekly warfarin dose	Published warfarin dosing algorithms / population-specific models	Cross-population/he out evaluation
Using machine learning to predict the bleeding risk for patients with cardiac valve replacement treated with warfarin in hospitalized (2024) [57]	Warfarin bleeding-risk prediction	Five ML models; XGBoost best-performing model	High bleeding risk during hospitalization (INR ≥ 4.5 surrogate)	Alternative ML models / conventional assessment	Internal development/valida
Predictors of warfarin therapeutic dose across thromboembolic indications: an exploratory early clinical evaluation study using machine learning-integrated explainable artificial intelligence and Bayesian models (2026) [58]	Warfarin dose optimization / XAI	Five ML models + SHAP + BART	Therapeutic warfarin dose across AF/stroke and DVT/PE indications	Alternative ML and Bayesian models	Internal evaluation; prospective validati recommended
A Post-Hoc Interpretable Ensemble Model to Feature Effect Analysis in Warfarin Dose Prediction for Chinese Patients (2022) [59]	Warfarin dose optimization / XAI	Interpretable ensemble + permutation importance + LIME + SHAP	Daily warfarin dose	Component/baseline predictive models	Internal evaluation
Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning (2022) [60]	Dynamic warfarin dose optimization	Reinforcement learning	Sequential dose optimization and INR response	Clinicians and alternative algorithms	Held-out test set
Warfarin Dose Management Using Offline Deep Reinforcement Learning (2025) [61]	Dynamic warfarin dose optimization	Batch-Constrained Q-learning (BCQ)	Cumulative dose until next INR / dose-policy prediction	XGBoost regression and LSTM baseline	Offline held-out evaluation
Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation (2024) [62]	Clinical risk prediction under OAC	Multi-Gate Mixture of Experts, Gradient Boosting and other ML models	Major bleeding, ischemic stroke, all-cause/CV death	CHA_2DS_2-VASc and HAS-BLED	Cross-validation
An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set (2021) [63]	Warfarin dose optimization	LightGBM ensemble framework with cross-over variables	Warfarin maintenance dose	Baseline models and IWPC model	5-fold cross-validati

Study	Domain	Model	Outcome	Comparator	Validation
Optimizing warfarin dosing in diabetic patients through BERT model and machine learning techniques (2025) [64]	Warfarin dose optimization	BERT-informed representation + RF, KNN, MLP, SVM, linear regression	Individualized warfarin dose in patients with diabetes	Linear regression and alternative ML models	Train/test evaluatio
A Digital Twin Strategy to Predict Thrombotic Recurrence in Antiphospholipid Syndrome Patients Treated with Direct Oral Anticoagulants vs. Vitamin K Antagonists Using Data from Real-World Populations (2025) [65]	Individualized OAC selection / thrombotic recurrence	Conditional GAN digital twin + causal DAG	Thrombotic recurrence and individualized counterfactual treatment effects	DOAC vs VKA	Synthetic-fidelity assessment / counterfactual simulation
Evaluation of supervised machine learning algorithms in predicting the poor anticoagulation control and stable weekly doses of warfarin (2023) [66]	Warfarin dose / anticoagulation-control prediction	C5.0, CART and CHAID	Poor anticoagulation control and stable weekly dose	Alternative tree-based algorithms	Internal evaluation
Determining the adjusted initial treatment dose of warfarin anticoagulant medicine using kernel-based support vector regression (2022) [67]	Warfarin initial-dose optimization	Kernel-based support vector regression	Adjusted initial warfarin dose	Conventional/predefined kernels	Internal evaluation
Unsupervised clustering approach to assess heterogeneity of treatment effects across patient phenotypes in randomized clinical trials (2025) [68]	Individualized OAC selection / HTE	Unsupervised phenotype clustering	Stroke/systemic embolism treatment-effect heterogeneity	Edoxaban vs warfarin	Randomized-trial secondary analysis
Predicting Range of Initial Warfarin Dose Based on Pharmacometabolomic and Genetic Inputs (2021) [69]	Warfarin initial-dose optimization	PLS + LDA/multinomial-logit ML + IniWarD	Day-7 INR response and initial dose range	Observed INR response / dose-response strata	Training/test split (120/40)

For warfarin dose optimization, several models reported strong performance within individual datasets. A support-vector model after mechanical heart-valve replacement reported R^2=0.98 and 93.83% of predictions within the prespecified ideal range [20]. An ensemble framework using 377 patients and 1,173 order events achieved a mean R^2 of 0.75 in five-fold cross-validation [63]. A pharmacometabolomic/pharmacogenetic model of initial warfarin response reported R^2=0.753 in training and 0.643 in testing, with 91% and 90% subgroup-classification accuracy, respectively [69]. Dynamic approaches extended beyond static dose prediction: a reinforcement-learning study reported an excellent-responder ratio of 80.8% under the learned policy compared with 41.6% under clinicians in the held-out test set [60], while an offline BCQ model reported 98.6% dose-policy prediction accuracy [61]. These retrospective and counterfactual metrics do not establish prospective clinical benefit or safety.

Clinical risk-prediction studies generally showed more modest discrimination. In 11,078 anticoagulated patients with atrial fibrillation, gradient boosting improved major-bleeding prediction among DOAC users compared with HAS-BLED (AUC 0.711 versus 0.586), whereas ischemic-stroke prediction was limited (AUC 0.606) in the presence of few outcome events [62]. Other DOAC studies reported moderate discrimination for bleeding or intracranial haemorrhage [25, 37, 46]. Improvements over conventional scores therefore did not necessarily indicate high absolute predictive accuracy.

Individualized treatment-selection studies were fewer and methodologically distinct. In a secondary analysis of ENGAGE AF–TIMI 48 (n = 14,062), phenotype clustering identified heterogeneity in the effect of higher-dose edoxaban versus warfarin on stroke or systemic embolism (interaction p = 0.03), with hazard ratios ranging from 0.72 to 1.01 across phenotypes [68]. A digital-twin analysis in antiphospholipid syndrome used real-world DOAC and VKA data to simulate counterfactual recurrence risk [65]. Other causal-ML studies estimated patient-level heterogeneity across oral-anticoagulant strategies [17, 19, 40]. These analyses require stronger causal assumptions than ordinary prognostic models and remain dependent on external and prospective validation.

Comparison of model families and conventional approaches

Across studies, AI-based models were commonly compared with conventional regression, established risk scores, pharmacogenetic algorithms, clinicians, or alternative machine-learning approaches rather than with a single universal comparator. Ensemble and tree-based models frequently achieved favorable or comparable performance, particularly in nonlinear dose-prediction and bleeding-risk tasks. Support-vector approaches demonstrated strong results in

selected warfarin dosing cohorts, and neural/deep-learning methods were most useful where longitudinal or high-dimensional input data were available. Reinforcement-learning methods offered a conceptually different approach by learning a treatment policy rather than a static prediction function. Causal and HTE methods addressed treatment selection rather than conventional prediction. Because model families were applied to different clinical tasks and datasets, no single architecture can be considered universally superior.

Generalizability and evidence gaps

Despite encouraging findings, several features limited generalizability. Many studies were retrospective, single-center, country-specific, or dependent on a single pharmacogenomic dataset. Independent external validation was less common than internal split-sample or cross-validation. Calibration, clinical utility, decision-curve analysis, and model updating were inconsistently reported. In dose-prediction studies, very high internal R^2 or PW20 values may reflect local model fit and do not guarantee transportability. In clinical-event studies, improvements in discrimination were sometimes statistically meaningful but modest in absolute terms. For reinforcement-learning and digital-twin studies, counterfactual policy performance remained dependent on modeling assumptions and could not substitute for prospective treatment assignment.

Methodological quality

Methodological appraisal identified the analysis/reporting domain as the principal source of risk-of-bias concern. Across the harmonized 54-study assessment, 13 studies were judged overall low risk, 13 had moderate risk/some concerns, and 28 were judged high/serious risk. Participants/data sources and predictor definitions were generally better reported than model-analysis procedures. Common analytical concerns included small effective sample size relative to model complexity, extensive feature/model selection, simple random train-test splits, synthetic oversampling, incomplete calibration reporting, lack of independent validation, and retrospective or counterfactual evaluation of treatment policies. Figure 2a shows study-level overall judgments and Fig. 2b summarizes harmonized domain-level proportions. The underlying study-level domain judgments are provided in Supplementary Data 1, Table S4.

Discussion

This systematic review synthesizes the current evidence for AI/ML in personalized oral anticoagulation management. Three findings were particularly consistent. First, warfarin dose optimization remains the most mature application, with models evolving from static pharmacogenomic prediction toward longitudinal recurrent networks and reinforcement-learning policies. Second, clinical risk prediction under active OAC therapy is expanding rapidly, especially for bleeding outcomes in DOAC-treated populations, but absolute discrimination is generally more modest than the strongest dose-prediction results. Third, causal ML, phenotype clustering, and digital-twin methods are beginning to shift the field from risk prediction toward individualized treatment selection. Across all three domains, however, methodological heterogeneity, limited external validation, and sparse prospective clinical evaluation remain major barriers to implementation.

The prominence of warfarin is clinically understandable because its dose-response relationship is influenced by genetic, demographic, dietary, medication, and longitudinal INR factors. Recent systematic evidence similarly indicates that ML-based warfarin models frequently outperform conventional approaches within individual datasets, while emphasizing that external validation and clinical utility remain inconsistent [6]. The present review extends that evidence by including dynamic and sequential decision-making studies. LSTM and reinforcement-learning approaches are particularly relevant because warfarin management is inherently longitudinal: the clinical decision at a given visit depends on previous doses, previous INR values, current INR, and evolving patient status rather than on a single baseline feature vector.

For DOACs and mixed OAC populations, the evidence was more strongly centered on risk prediction than dose calculation. Large real-world studies have shown that ML can outperform established bleeding scores, but the gains are generally moderate and outcome-specific [7, 24–25, 37]. This distinction is clinically important. A statistically higher AUC does not automatically imply sufficient calibration, net benefit, or decision support for individual patients. Similarly, residual thrombotic-risk models demonstrate that clinically important events still occur despite active anticoagulation, but prediction of rare outcomes such as ischemic stroke remains difficult when event counts are limited [24].

The treatment-selection literature represents a conceptually different direction. Rather than estimating prognosis under a single therapy, causal ML and HTE methods estimate how expected benefit and harm vary across treatment strategies. Recent pooled and trial-based work comparing DOACs with warfarin demonstrates the potential of this approach [26]. Digital-twin methods also provide a framework for simulating counterfactual treatment scenarios, but their credibility depends on the quality of the observed data, control of confounding, model fidelity, and appropriate validation [65]. These methods should therefore complement rather than replace randomized and well-designed comparative evidence.

The immediate clinical value of AI is likely to be greatest as an adjunct to, rather than a replacement for, clinician-led anticoagulation management. For warfarin, AI may support initial or maintenance dose selection, identification of patients at risk of poor INR control, and dynamic dose adjustment using longitudinal data. For DOACs, clinically plausible applications include individualized bleeding-risk estimation, identification of residual thrombotic risk, and treatment-selection support when several approved anticoagulant options are available. Current clinical evidence on dose-specific DOAC efficacy and safety reinforces the importance of selecting the appropriate drug and dose for an individual patient [14].

Explainability is especially important in anticoagulation because model outputs may influence decisions with immediate bleeding or thrombotic consequences. The 2026 XAI warfarin study illustrates both the promise and limitation of explainability: SHAP and permutation importance clarified the contribution of VKORC1, CYP2C9, CYP4F2, drug interactions, and clinical variables, but overall discrimination remained modest [39]. Explainability therefore cannot compensate for weak predictive validity; instead, it should be viewed as an additional requirement for transparency, clinical review, and identification

of model failure modes. Broader reviews of AI in thrombosis and AF similarly emphasize interpretability, workflow integration, governance, and clinician trust as prerequisites for clinical translation [3–5, 15].

Several methodological concerns recur across this literature. Complex ML architectures were often developed using relatively small or highly selected datasets, creating substantial risk of overfitting. Internal split-sample validation was common, whereas independent external or prospective validation was less frequent. Calibration was reported inconsistently, despite its importance when predicted probabilities are used to guide treatment. Outcome heterogeneity also complicated evidence synthesis: warfarin studies targeted initial dose, maintenance dose, INR, TTR, or dosing-policy agreement, whereas risk models targeted bleeding, intracranial hemorrhage, stroke, thrombosis, or composite outcomes. Consequently, apparent differences in model performance frequently reflect different tasks rather than differences in algorithmic quality.

The distinction between prediction and causal treatment-effect estimation is also methodologically important. Prediction models can be well calibrated for an outcome without identifying which treatment is best for an individual patient. Conversely, treatment-effect models require assumptions regarding exchangeability, treatment assignment, positivity, confounding, and consistency that do not apply to ordinary prognostic prediction. For this reason, the present review used design-appropriate risk-of-bias frameworks rather than applying a single prediction-model tool to all included studies.

This review has several strengths. Eligibility was deliberately restricted to adults actively receiving oral anticoagulants and required AI/ML to have a direct role in dose optimization, clinically actionable risk prediction, or treatment selection. This prevented dilution of the evidence base with AI studies focused only on AF detection, prescribing behavior, medication adherence, workflow automation, or incidental anticoagulant exposure. Screening followed a prespecified PICO-based framework, and the final synthesis separates conventional prediction, sequential treatment optimization, and causal/HTE applications.

This review has several limitations. Searches were restricted to English-language reports and to PubMed, Europe PMC and Scopus; overlap between PubMed and Europe PMC and the absence of Embase and Web of Science may have reduced retrieval breadth and could have introduced selection bias. The evidence was heterogeneous in population, model architecture, predictors, validation, outcome definitions and performance metrics, precluding a clinically meaningful meta-analysis. Several publications reused major datasets, including IWPC and randomized-trial cohorts, so the 54 included reports should not be interpreted as 54 independent patient populations. Many studies reported internal validation or surrogate and counterfactual outcomes, and calibration and clinical-utility analyses were inconsistent. The available documentation does not establish fully independent duplicate data extraction. Publication or selective-reporting bias cannot be excluded, and no formal certainty-of-evidence grading was conducted. Finally, the appraisal dataset contains harmonized domain judgments rather than full signaling-question responses, which limits independent reproduction of every appraisal judgment. These limitations constrain the certainty and completeness of the evidence and support cautious interpretation of apparent performance advantages until externally and prospectively validated.

Future research should prioritize prospective multicenter validation, transparent reporting of calibration and missing-data handling, prespecified model evaluation, and evaluation against clinically meaningful comparators. Reinforcement-learning policies should be tested prospectively before use for autonomous dose adjustment. Causal and HTE models should be validated across independent datasets and, where feasible, against randomized evidence. Explainable AI should be evaluated not only for visual interpretability but also for whether explanations improve clinician understanding, decision consistency, patient outcomes, and safety. Ultimately, the field should move from repeated retrospective model development toward rigorously evaluated clinical decision-support systems integrated with anticoagulation workflows.

AI and ML have expanded personalized oral anticoagulation management from static warfarin dose prediction toward longitudinal dose optimization, prediction of bleeding and thromboembolic risk during active treatment, and individualized treatment selection. Several model families demonstrate favorable performance within individual studies, particularly when clinical, pharmacogenomic, and longitudinal data are integrated. However, the evidence base remains limited by methodological heterogeneity, concentration of risk-of-bias concerns in the analysis domain, inconsistent calibration and reporting, limited independent validation, and sparse prospective evaluation of patient-important outcomes. AI-guided anticoagulation should therefore be considered a promising clinical decision-support strategy that requires further prospective multicenter validation before routine implementation.

Methods

Study design registration and reporting

This systematic review was reported in accordance with PRISMA 2020 [9]. The review protocol was registered in PROSPERO (CRD420261502629), and the registration record is available through PROSPERO.

Search strategy

PubMed, Europe PMC and Scopus were searched from database inception to 15 August 2026. The database set was selected to capture biomedical literature together with broader interdisciplinary and technical literature relevant to artificial intelligence and clinical care. Database-specific title, abstract and keyword syntax was used to combine concepts for artificial intelligence or machine learning, oral anticoagulants, and anticoagulation-management outcomes. Searches were limited to English-language reports. The exact strategies and pre-deduplication result counts are reproduced in the Supplementary Information. Because Europe PMC overlaps substantially with PubMed, and Embase and Web of Science were not searched, the search may not capture all eligible studies; this limitation was considered when interpreting the completeness of the evidence base.

Eligibility criteria

Study eligibility was predefined using a PICOS framework. Eligible populations were adults actively receiving oral anticoagulant therapy, including warfarin or other vitamin K antagonists (VKAs) and DOACs such as apixaban, rivaroxaban, dabigatran, and edoxaban. Eligible indications included AF, venous thromboembolism, mechanical heart valves, antiphospholipid syndrome, and other indications for oral anticoagulation. Studies were excluded when anticoagulation was incidental, historical, restricted to parenteral agents or antiplatelets, or when the target population was not actively receiving OAC therapy. The intervention of interest was a substantive patient-level AI/ML model directly used to predict, select, adjust, or optimize oral anticoagulant therapy or a clinically actionable anticoagulation-related outcome. Eligible comparators included clinician-guided management, conventional regression models, clinical scores, established dosing algorithms, alternative AI/ML models, alternative anticoagulants or doses, or no comparator in eligible model-development studies. Eligible outcomes included dose and INR prediction, time in therapeutic range (TTR), over- or under-anticoagulation, major or clinically relevant bleeding, gastrointestinal or intracranial hemorrhage, ischemic stroke, systemic embolism, recurrent or residual thrombosis, and individualized treatment or dose selection. Original patient-level model development, validation, trial secondary analyses, observational cohorts, registries, EHR/claims studies, and clinical implementation studies were eligible. Reviews, protocols without results, pharmacovigilance-only studies, molecular/QSAR studies, simulation-only studies without real patients, and hypothetical vignette studies were excluded.

Study selection

All records were imported into Rayyan [10], and 405 duplicates were removed. Of 579 unique records screened, 507 were excluded during record-level title/abstract screening and final eligibility verification before report retrieval, leaving 72 reports sought for retrieval. One report could not be retrieved, and 71 full texts were assessed. Screening was conducted independently by AA and MM. Disagreements were resolved through discussion, with RBH acting as adjudicator when required. Supplementary Data 1, Table S2 documents the 72 reports sought for retrieval and their full-text eligibility decisions.

Data extraction

AA and MM contributed to data extraction using a standardized Microsoft Excel form. Extracted variables included publication year, study design, data source, sample size, population and indication, oral anticoagulant, model purpose and architecture, comparator, outcome, validation strategy, performance metrics, explainability, clinical utility, main findings and reported limitations. Relevant results were retained when studies evaluated several eligible models, and unreported information was recorded as NR rather than inferred. The available documentation does not establish fully independent duplicate extraction; this is acknowledged as a review-process limitation.

Risk of bias assessment

Methodological quality was evaluated using a structured, design-appropriate, framework-informed appraisal. Prediction-model studies were examined across participants or data sources, predictors, outcomes, and analysis or reporting with reference to PROBAST + AI [11]. Observational causal-ML studies were considered using ROBINS-I principles [13], and randomized-trial secondary treatment-effect analyses using RoB 2 principles [12], with additional attention to post hoc model selection and treatment-effect estimation. Because these frameworks address different study designs and are not directly interchangeable, harmonized low, moderate/some-concerns, and high/serious-concern categories were used only for descriptive visualization and cross-study presentation. Tool-specific methodological concerns were retained in interpretation, and the harmonized categories were not treated as a pooled quality score or as equivalent judgments across frameworks. The supplied appraisal dataset reports domain-level judgments rather than the full item-level signaling-question record; findings should therefore be interpreted as a structured methodological appraisal rather than a complete reproducible application of each instrument.

Data synthesis

A narrative synthesis was structured by clinical application: dose optimization or anticoagulation control, clinical risk prediction during active oral-anticoagulant therapy, and individualized treatment selection or heterogeneous treatment effects. Model characteristics, validation strategies, comparators, performance measures, explainability and clinical utility were summarized descriptively. Quantitative pooling was not undertaken because populations, treatment targets, outcome definitions, validation strategies and performance metrics were not sufficiently comparable, and combining prediction, sequential-policy and causal treatment-effect estimates would not yield a clinically interpretable summary measure. Results from these methodological paradigms were therefore interpreted separately. No formal certainty-of-evidence grading or statistical assessment of reporting bias was performed because the evidence base comprised heterogeneous model-development, validation, sequential-policy and treatment-effect studies that did not share a common estimand or synthesis framework; this limitation was explicitly considered in interpretation.

Ethics

Ethics approval

and participant consent were not required because this systematic review used data already reported in published studies and did not access identifiable individual-participant data.

Clinical trial number

Clinical trial number: not applicable.

Data availability

All data analysed during this systematic review are included in the article and its supplementary materials. The supplementary files contain the study characteristics, screening and eligibility information, and structured methodological appraisal data supporting the findings of this review.

Code availability

No custom computer code was used as a central analytic component of this systematic review.

Declarations

Competing interests

All authors declare no financial or non-financial competing interests.

Funding

This research did not receive funding.

Author Contribution

A.A. and M.M. conducted the literature screening and eligibility assessment and contributed to data extraction, analysis, interpretation, and manuscript preparation. A.A. led manuscript drafting and project coordination. R.B.H. provided academic supervision, adjudication of disagreements when required, manuscript refinement, and critical review. N.A.H.J. contributed to manuscript review, refinement, and critical revision. F.B.A.G. contributed to manuscript review, clinical refinement, and critical revision. All authors reviewed and approved the final manuscript.

Acknowledgement

ChatGPT (OpenAI) was used during manuscript preparation to assist with language refinement, clarity, organization, and formatting. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the scientific content and final manuscript.

Data Availability

All data analysed during this systematic review are included in the published article and its supplementary materials. The supplementary files contain the study characteristics, screening and eligibility information, and structured methodological appraisal data supporting the findings of this review.

References

1. Cross B, Turner RM, Zhang JE, Pirmohamed M. Being precise with anticoagulation to reduce adverse drug reactions: are we there yet? Pharmacogenomics J. 2024;24:7. 10.1038/s41397-024-00329-y.
2. Castro-Vidal ZA, Gandapur A, Prakash R et al. Anticoagulation strategies for stroke prevention in atrial fibrillation: a comprehensive review of current literature. Ann Med Surg (Lond). 2025.
3. Karakasis P, Theofilis P, Sagris M, et al. Artificial Intelligence in Atrial Fibrillation: From Early Detection to Precision Therapy. J Clin Med. 2025;14:2627. 10.3390/jcm14082627.
4. Antoun I, Abdelrazik A, Eldesouky M, et al. Artificial intelligence in atrial fibrillation: emerging applications, research directions and ethical considerations. Front Cardiovasc Med. 2025;12:1596574.
5. Tica O, Champsi A, Duan J, Tica O. Artificial Intelligence in the Diagnosis and Management of Atrial Fibrillation. Diagnostics. 2025;15:2561.
6. Fülöp P, Tóth Š, Porubän T, Fülöpová Z, Borovská A, Dvorožňáková M. Machine Learning for Warfarin Therapy: A Systematic Review. Pharmaceuticals. 2025;18:1544.
7. Teo WZY, Wong MWY, Lim FJ, et al. Machine Learning Models for Predicting Bleeding Risk in Anticoagulated Patients with Atrial Fibrillation and Venous Thromboembolism: A Comparative Evidence Synthesis. J Clin Med. 2026;15:2370.
8. Woodman RJ, Ng HS, Mangoni AA. Clinical phenotypes of atrial fibrillation: a review of machine learning applications in personalized treatment. JRSM Cardiovasc Dis. 2025;14:20480040251399275.
9. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
10. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5:210.
11. Moons KGM, Damen JAA, Kaul T, et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505.
12. Sterne JAC, Savović J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. 2019.
13. Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
14. Oh SH, Cheon S, Choi SY, Kim YS, Choi HG, Chung JE. Effectiveness and Safety of Dose-Specific DOACs in Patients With Atrial Fibrillation: A Systematic Review and Network Meta-Analysis. Cardiovasc Ther. 2025;2025:9923772.
15. Kuan YKI, Kok YJ, Liu NSH et al. Artificial intelligence in clinical thrombosis and hemostasis: A review. Res Pract Thromb Haemost. 2025.

16. Wang Z, Poon J, Yang J, Poon S. Warfarin dose estimation on high-dimensional and incomplete data. Proceedings of the Annual Hawaii International Conference on System Sciences. 2021;2020:3455–3463.
17. Ngufor C, Yao X, Inselman JW, Ross JS, Dhruva SS, Graham DJ, et al. Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method. Am Heart J. 2023;260:124–40. 10.1016/j.ahj.2023.02.015.
18. Truda G, Marais P. Evaluating warfarin dosing models on multiple datasets with a novel software framework and evolutionary optimisation. J Biomed Inform. 2021;113:103634. 10.1016/j.jbi.2020.103634.
19. Xu Y, Bechler K, Callahan A, Shah N. Principled estimation and evaluation of treatment effect heterogeneity: A case study application to dabigatran for patients with atrial fibrillation. J Biomed Inform. 2023;143:104420. 10.1016/j.jbi.2023.104420.
20. Guo B, Chen C, Jia J, Zheng J, Song Y, Liu T, et al. A Prediction Model of Stable Warfarin Doses in Patients After Mechanical Heart Valve Replacement Based on a Machine Learning Algorithm. Rev Cardiovasc Med. 2025;26(6):33425. 10.31083/RCM33425.
21. Khianchainat K, Kanjanawattana SA. Comparative Analysis of Machine Learning Techniques for Therapeutic Warfarin Dose Prediction. ACM International Conference Proceeding Series. 2024:59–63. 10.1145/3702468.3702478
22. Gan TRX, Tan LWJ, Egermark M, Truong ATL, Kumar K, Tan SB, et al. AI-assisted warfarin dose optimisation with CURATE.AI for clinical impact: Retrospective data analysis. Bioeng translational Med. 2025;10(3):e10757. 10.1002/btm2.10757.
23. Asiimwe IG, Blockman M, Cohen K, Cupido C, Hutchinson C, Jacobson B, et al. Stable warfarin dose prediction in sub-Saharan African patients: A machine-learning approach and external validation of a clinical dose-initiation algorithm. CPT: pharmacometrics Syst Pharmacol. 2022;11(1):20–9. 10.1002/psp4.12740.
24. Liu Y, Chen Y, Olier I, Ortega-Martorell S, Huang B, Ishiguchi H, et al. Residual risk prediction in anticoagulated patients with atrial fibrillation using machine learning: A report from the GLORIA-AF registry phase II/III. Eur J Clin Invest. 2025;55(3):e14371. 10.1111/eci.14371.
25. Chaudhary R, Nourelahi M, Thoma FW, Gellad WF, Lo-Ciganic WH, Chaudhary R, et al. Machine Learning Predicts Bleeding Risk in Atrial Fibrillation Patients on Direct Oral Anticoagulant. Am J Cardiol. 2025;244:58–66. 10.1016/j.amjcard.2025.02.030.
26. Al Said S, Bellavia A, Braunwald E, Hylek EM, Palazzolo MG, Hong H, et al. Heterogeneity of Treatment Effect of DOACs vs Warfarin in Atrial Fibrillation. JACC Adv. 2026;5(7):102915. 10.1016/j.jacadv.2026.102915.
27. Dryden L, Song J, Valenzano TJ, Yang Z, Debnath M, Lin R, et al. Evaluation of Machine Learning Approaches for Predicting Warfarin Discharge Dose in Cardiac Surgery Patients: Retrospective Algorithm Development and Validation Study. JMIR cardio. 2023;7:e47262. 10.2196/47262.
28. Ishii M, Kaikita K, Yasuda S, Akao M, Ako J, Matoba T, et al. Risk prediction score for clinical outcome in atrial fibrillation and stable coronary artery disease. Open heart. 2023;10(1). 10.1136/openhrt-2023-002292.
29. Wang M, Qian Y, Yang Y, Chen H, Rao WF. Improved stacking ensemble learning based on feature selection to accurately predict warfarin dose. Front Cardiovasc Med. 2023;10:1320938. 10.3389/fcvm.2023.1320938.
30. Ratajczak F, Hohendanner F, Haack S, Boldt L, Parwani AS, Blaschke F, et al. Intracardiac thrombus formations despite continuous oral anticoagulation in atrial fibrillation patients undergoing catheter ablation procedures: pilot development of a machine learning prediction model. Front Cardiovasc Med. 2026;13:1707003. 10.3389/fcvm.2026.1707003.
31. Roche-Lima A, Roman-Santiago A, Feliu-Maldonado R, Rodriguez-Maldonado J, Nieves-Rodriguez BG, Carrasquillo-Carrion K, et al. Machine Learning Algorithm for Predicting Warfarin Dose in Caribbean Hispanics Using Pharmacogenetic Data. Front Pharmacol. 2019;10:1550. 10.3389/fphar.2019.01550.
32. Steiner HE, Giles JB, Patterson HK, Feng J, El Rouby N, Claudio K, et al. Machine Learning for Prediction of Stable Warfarin Dose in US Latinos and Latin Americans. Front Pharmacol. 2021;12:749786. 10.3389/fphar.2021.749786.
33. Dai MF, Li SY, Zhang JF, Wang BY, Zhou L, Yu F, et al. Warfarin anticoagulation management during the COVID-19 pandemic: The role of internet clinic and machine learning. Front Pharmacol. 2022;13:933156. 10.3389/fphar.2022.933156.
34. Li Y, Xu N, Song L, Liu Y, Pei Q, Kuang Y, et al. AI-WAR: a novel warfarin management software with a bidirectional LSTM dosing model improves time in therapeutic range. Front Pharmacol. 2026;17:1750503. 10.3389/fphar.2026.1750503.
35. Huang YC, Cheng YC, Jhou MJ, Chen M, Lu CJ. Integrated Machine Learning Decision Tree Model for Risk Evaluation in Patients with Non-Valvular Atrial Fibrillation When Taking Different Doses of Dabigatran. Int J Environ Res Public Health. 2023;20(3). 10.3390/ijerph20032359.
36. Huang YC, Cheng YC, Jhou MJ, Chen M, Lu CJ. Important Risk Factors in Patients with Nonvalvular Atrial Fibrillation Taking Dabigatran Using Integrated Machine Learning Scheme-A Post Hoc Analysis. J personalized Med. 2022;12(5). 10.3390/jpm12050756.
37. Liu Y, Li L, Li J, Liu H, Geru A, Wang Y, et al. Development and Validation of a Predictive Model for Intracranial Haemorrhage in Patients on Direct Oral Anticoagulants. Clin Appl thrombosis/hemostasis: official J Int Acad Clin Appl Thrombosis/Hemostasis. 2024;30:10760296241271338. 10.1177/10760296241271338.
38. Xue L, He, Shan S, Qin RK, Qiong, Ding Y, Liu L, et al. Machine learning guided prediction of warfarin blood levels for personalized medicine based on clinical longitudinal data from cardiac surgery patients: a prospective observational study. Int J Surg (London England). 2024;110(10):6528–40. 10.1097/js9.0000000000001734.
39. Sridharan K, Sivaramakrishnan G. Explainable Artificial Intelligence Unravels the Possible Distinct Roles of VKORC1 and CYP2C9 in Predicting Warfarin Anticoagulation Control. Medical sciences (Basel, Switzerland). 2026;14(1). 10.3390/medsci14010156
40. Meid AD, Wirbka L, Groll A, Haefeli WE. Med Decis making: Int J Soc Med Decis Mak. 2022;42(5):587–98. 10.1177/0272989X211064604. Can Machine Learning from Real-World Data Support Drug Treatment Decisions? A Prediction Modeling Case for Direct Oral Anticoagulants.

41. Kim JH, Yee J, Chang BC, Gwak HS. Gene Polymorphisms of the Renin-Angiotensin System and Bleeding Complications of Warfarin: Genetic-Based Machine Learning Models. Pharmaceuticals (Basel Switzerland). 2021;14(8). 10.3390/ph14080824.
42. Kim J, Jang I. Predictors of bleeding event among elderly patients with mechanical valve replacement using random forest model: A retrospective study. Medicine. 2021;100(19):e25875. 10.1097/MD.0000000000025875.
43. Tao Y, Jiang B, Xue L, Xie C, Zhang Y. Evolutionary synthetic oversampling technique and cocktail ensemble model for warfarin dose prediction with imbalanced data. Neural Comput Appl. 2021;33(17):11203–21. 10.1007/s00521-020-05568-1.
44. Gu ZC, Huang SR, Dong L, Zhou Q, Wang J, Fu B, et al. Cardiovasc Drugs Ther. 2022;36(5):879–89. 10.1007/s10557-021-07191-1. An Adapted Neural-Fuzzy Inference System Model Using Preprocessed Balance Data to Improve the Predictive Accuracy of Warfarin Maintenance Dosing in Patients After Heart Valve Replacement.
45. Han H, Zhang J, Wang X, Ge W, Qu JZ. Development of an interpretable machine learning model to predict short-term bleeding risk in patients receiving dual antithrombotic therapy following cardiac surgery. Int J Clin Pharm. 2026;48(2):513–23. 10.1007/s11096-025-02016-9.
46. Chen C, Yin C, Wang Y, Zeng J, Wang S, Bao Y, et al. XGBoost-based machine learning test improves the accuracy of hemorrhage prediction among geriatric patients with long-term administration of rivaroxaban. BMC Geriatr. 2023;23(1):418. 10.1186/s12877-023-04049-z.
47. Tao Y, Zhang Y, Jiang B. DBCSMOTE: a clustering-based oversampling technique for data-imbalanced warfarin dose prediction. BMC Med Genom. 2020;13:152. 10.1186/s12920-020-00781-2.
48. Nwanosike EM, Sunter W, Ansari MA, Merchant HA, Conway B, Hasan SS. A Real-World Exploration into Clinical Outcomes of Direct Oral Anticoagulant Dosing Regimens in Morbidly Obese Patients Using Data-Driven Approaches. American journal of cardiovascular drugs: drugs, devices, and other interventions. 2023;23(3):287–99. 10.1007/s40256-023-00569-6
49. Li Q, Wang J, Tao H, Zhou Q, Chen J, Fu B, et al. The Prediction Model of Warfarin Individual Maintenance Dose for Patients Undergoing Heart Valve Replacement, Based on the Back Propagation Neural Network. Clin Drug Investig. 2020;40(1):41–53. 10.1007/s40261-019-00850-0.
50. Ma W, Li H, Dong L, Zhou Q, Fu B, Hou JL, et al. Warfarin maintenance dose prediction for Chinese after heart valve replacement by a feedforward neural network with equal stratified sampling. Sci Rep. 2021;11(1):13778. 10.1038/s41598-021-93317-2.
51. Lee H, Kim HJ, Chang HW, Kim DJ, Mo J, Kim JE. Development of a system to support warfarin dose decisions using deep neural networks. Sci Rep. 2021;11(1):14745. 10.1038/s41598-021-94305-2.
52. Choi H, Kang HJ, Ahn I, Gwon H, Kim Y, Seo H, et al. Machine learning models to predict the warfarin discharge dosage using clinical information of inpatients from South Korea. Sci Rep. 2023;13(1):22461. 10.1038/s41598-023-49831-6.
53. Petch J, Nelson W, Wu M, Ghassemi M, Benz A, Fatemi M, et al. Optimizing warfarin dosing for patients with atrial fibrillation using machine learning. Sci Rep. 2024;14(1):4516. 10.1038/s41598-024-55110-9.
54. Tezcan H, Tunçez A, Özen Y, Yalçın MU, Gürses KM. Retrospective modeling study of ChatGPT (GPT 4) warfarin dose adjustment in patients with INRs outside the therapeutic range. Digit health. 2026;12:20552076251412985. 10.1177/20552076251412985.
55. Kuang Y, Liu Y, Pei Q, Ning X, Zou Y, Liu L, et al. Long Short-Term Memory Network for Development and Simulation of Warfarin Dosing Model Based on Time Series Anticoagulant Data. Front Cardiovasc Med. 2022;9. 10.3389/fcvm.2022.881111.
56. Jahmunah V, Chen S, Oh SL, Acharya UR, Chowbay B. Automated warfarin dose prediction for Asian, American, and Caucasian populations using a deep neural network. Comput Biol Med. 2023;153:106548. 10.1016/j.compbiomed.2023.106548.
57. Hu Y, Zhang X, Wei M, Yang T, Chen J, Wu X, et al. Using machine learning to predict the bleeding risk for patients with cardiac valve replacement treated with warfarin in hospitalized. Pharmacoepidemiol Drug Saf. 2024;33(2):e5756. 10.1002/pds.5756.
58. Sridharan K, Sivaramakrishnan G. Predictors of warfarin therapeutic dose across thromboembolic indications: an exploratory early clinical evaluation study using machine learning-integrated explainable artificial intelligence and Bayesian models. Expert Rev Clin Pharmacol. 2026;19(6):587–600. 10.1080/17512433.2026.2682286.
59. Zhang Y, Xie C, Xue L, Tao Y, Yue G, Jiang B. A Post-Hoc Interpretable Ensemble Model to Feature Effect Analysis in Warfarin Dose Prediction for Chinese Patients. IEEE J Biomedical Health Inf. 2022;26(2):840–51. 10.1109/JBHI.2021.3092170.
60. Zeng J, Shao J, Lin S, Zhang H, Su X, Lian X, et al. Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning. J Am Med Inf Association: JAMIA. 2022;29(10):1722–32. 10.1093/jamia/ocac088.
61. Ji H, Gill M, Draper EW, Liedl DA, Hodge DO, Houghton DE, et al. Warfarin Dose Management Using Offline Deep Reinforcement Learning. IEEE J biomedical health Inf. 2025;29(7):5219–28. 10.1109/JBHI.2025.3545384.
62. Bernardini A, Bindini L, Antonucci E, Berteotti M, Giusti B, Testa S, et al. Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation. Int J Cardiol. 2024;407:132088. 10.1016/j.ijcard.2024.132088.
63. Liu Y, Chen J, You Y, Xu A, Li P, Wang Y, et al. An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set. Comput Biol Med. 2021;131:104242. 10.1016/j.compbiomed.2021.104242.
64. Ghafourian MS, Tarkiani S, Ghajar M, Chavooshi M, Khormaei H, Ramezani A. Optimizing warfarin dosing in diabetic patients through BERT model and machine learning techniques. Comput Biol Med. 2025;186:109755. 10.1016/j.compbiomed.2025.109755.
65. Casado-Suela MÁ, Torres-Macho J, Izquierdo-Martínez A, Ancos-Aracil CL, Ferreira-Burguillos L, Madroñal-Cerezo E, et al. A Digital Twin Strategy to Predict Thrombotic Recurrence in Antiphospholipid Syndrome Patients Treated with Direct Oral Anticoagulants vs. Vitamin K Antagonists Using Data from Real-World Populations. J Clin Med. 2025;14(16). 10.3390/jcm14165716.
66. Sridharan K, Ramanathan M, Al Banna R. Evaluation of supervised machine learning algorithms in predicting the poor anticoagulation control and stable weekly doses of warfarin. Int J Clin Pharm. 2023;45(1):79–87. 10.1007/s11096-022-01471-y.

67. Maghsoudi R, Mirzarezaee M, Sadeghi M, Nadjar-Araabi B. Determining the adjusted initial treatment dose of warfarin anticoagulant medicine using kernel-based support vector regression. Comput Methods Programs Biomed. 2022;214:106589. 10.1016/j.cmpb.2021.106589.
68. Bellavia A, Ran X, Zimerman A, Antman EM, Giugliano RP, Morrow DA, et al. Unsupervised clustering approach to assess heterogeneity of treatment effects across patient phenotypes in randomized clinical trials. Contemp Clin Trials. 2025;148:107778. 10.1016/j.cct.2024.107778.
69. Huang Q, Cao L, Luo N, Qian H, Wei M, Xue L, et al. Predicting Range of Initial Warfarin Dose Based on Pharmacometabolomic and Genetic Inputs. Clin Pharmacol Ther. 2021;110(6):1585–94. 10.1002/cpt.2407.

Figures

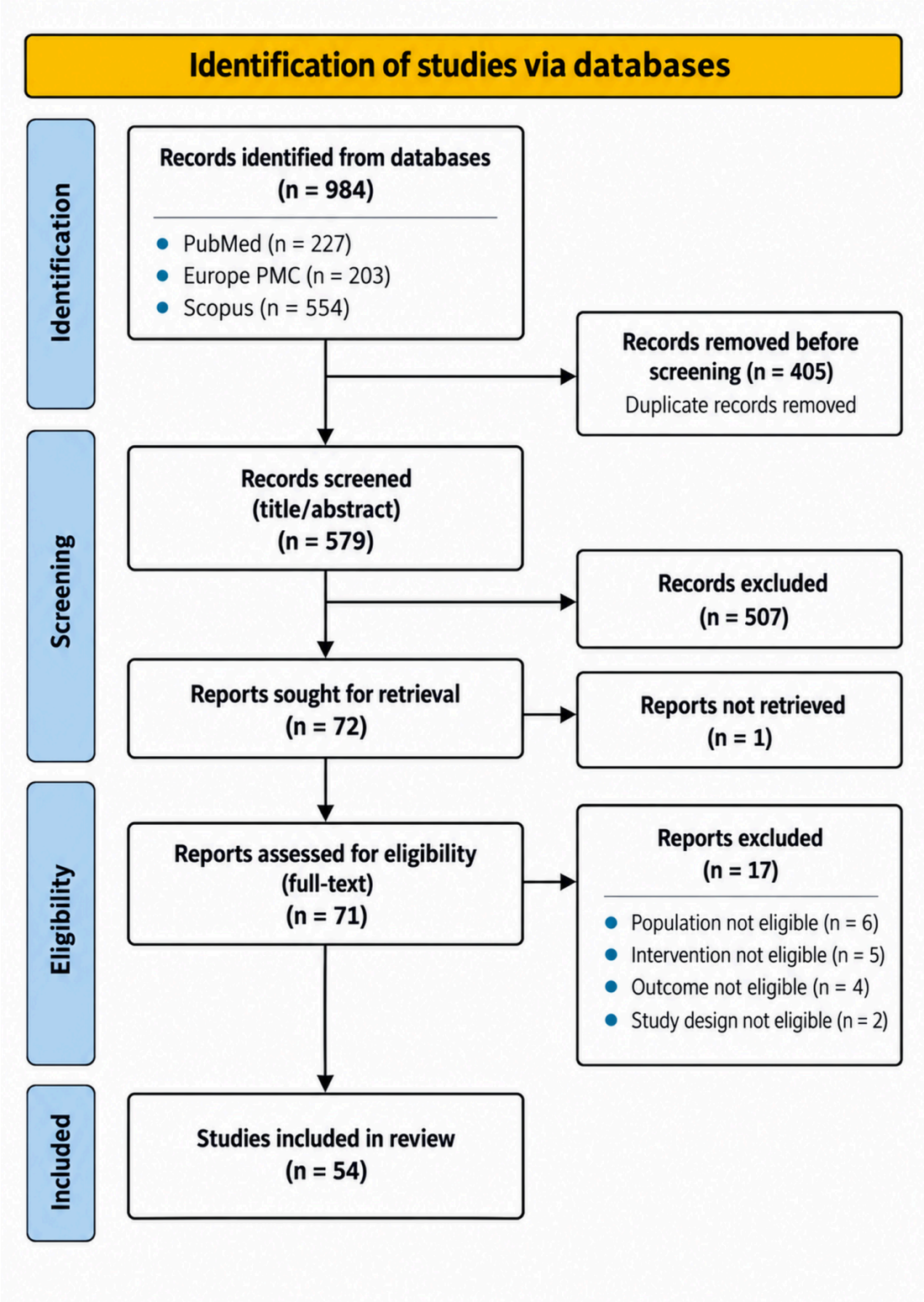
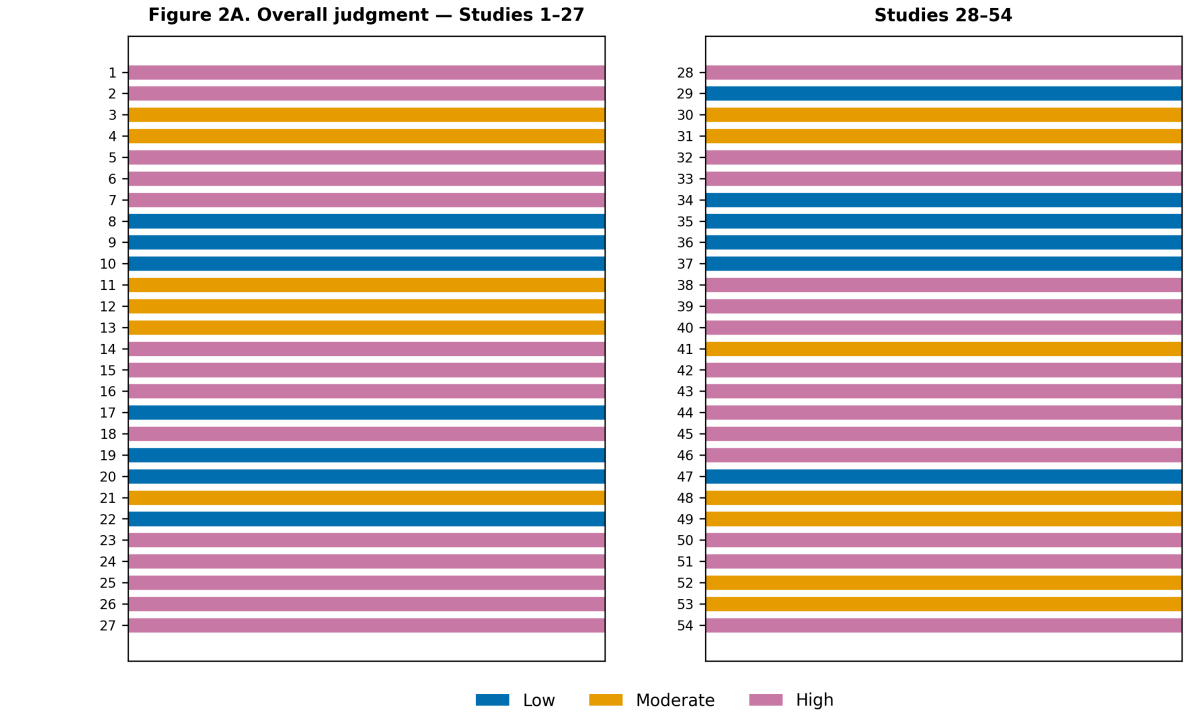


Figure 1

PRISMA 2020 flow diagram of study selection. The searches identified 984 records from PubMed, Europe PMC, and Scopus; 405 duplicates were removed before screening; 579 records were screened; 507 records were excluded before report retrieval; 72 reports were sought; one report was not retrieved; 71 full texts were assessed; 17 reports were excluded with reasons; and 54 studies were included.

Figure 2. Structured methodological appraisal of included studies



Study numbering follows Supplementary Data 1, Table S4.

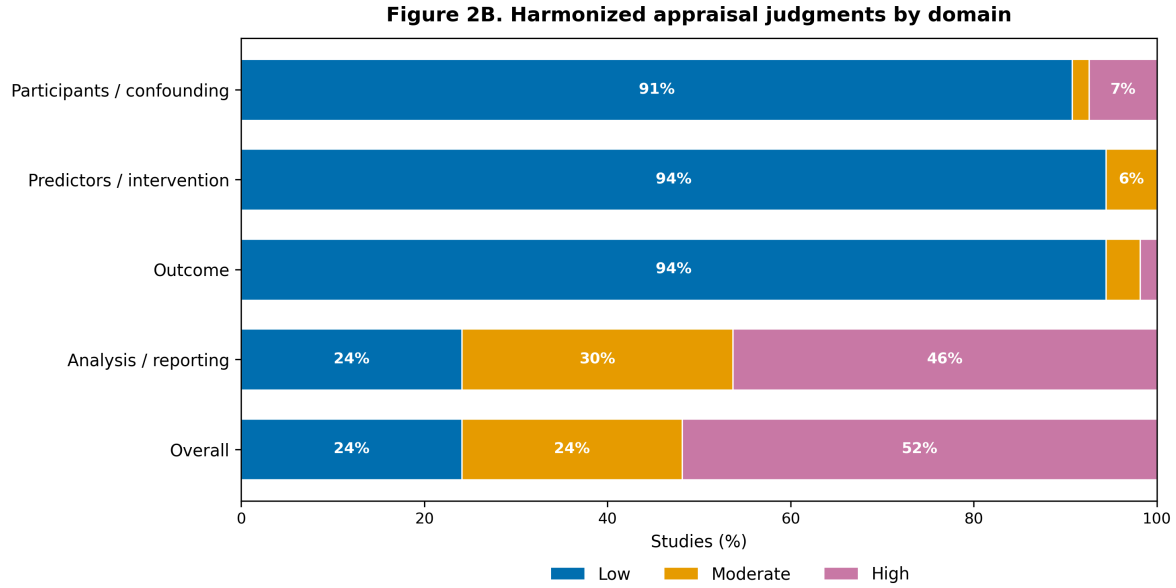


Figure 2

Structured methodological appraisal. a) Overall judgments for 54 included studies; study numbering follows Supplementary Data 1, Table S4. b) Harmonized domain-level proportions. Blue indicates low concern, amber moderate concern, and magenta high concern. The harmonized display does not make the source frameworks interchangeable.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 05SupplementaryData1.xlsx
- SupplementaryInformationSearchStrategiesREVISED.pdf
- PRISMA2020ChecklistScientificReports.pdf