



Article

Nonlinear Machine Learning in Warfarin Dose Prediction: Insights from Contemporary Modelling Studies

Fengying Zhang ^{1,†}, Yan Liu ^{2,†}, Weijie Ma ¹ , Shengming Zhao ¹, Jin Chen ^{1,*} and Zhichun Gu ^{3,4,*} 

¹ Department of Evidence-Based Medicine and Clinical Epidemiology, West China Hospital, Sichuan University, Chengdu 610041, China; xuanyoufenglaiyi@163.com (F.Z.); sxszmwj@163.com (W.M.); shengmin39@163.com (S.Z.)

² Department of Clinical Pharmacy, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200092, China; liuyan03@xinhua.com.cn

³ Department of Pharmacy, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

⁴ Shanghai Anticoagulation Pharmacist Alliance, Shanghai Pharmaceutical Association, Shanghai 200040, China

* Correspondence: ebm_chenjin@126.com (J.C.); guzhichun213@163.com (Z.G.)

† These authors contributed equally to this work.

Abstract: Objective: This study aimed to systematically assess the characteristics and risk of bias of previous studies that have investigated nonlinear machine learning algorithms for warfarin dose prediction. **Methods:** We systematically searched PubMed, Embase, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI), China Biology Medicine (CBM), China Science and Technology Journal Database (VIP), and Wanfang Database up to March 2022. We assessed the general characteristics of the included studies with respect to the participants, predictors, model development, and model evaluation. The methodological quality of the studies was determined, and the risk of bias was evaluated using the Prediction model Risk of Bias Assessment Tool (PROBAST). **Results:** From a total of 8996 studies, 23 were assessed in this study, of which 23 (100%) were retrospective, and 11 studies focused on the Asian population. The most common demographic and clinical predictors were age (21/23, 91%), weight (17/23, 74%), height (12/23, 52%), and amiodarone combination (11/23, 48%), while CYP2C9 (14/23, 61%), VKORC1 (14/23, 61%), and CYP4F2 (5/23, 22%) were the most common genetic predictors. Of the included studies, the MAE ranged from 1.47 to 10.86 mg/week in model development studies, from 2.42 to 5.18 mg/week in model development with external validation (same data) studies, from 12.07 to 17.59 mg/week in model development with external validation (another data) studies, and from 4.40 to 4.84 mg/week in model external validation studies. All studies were evaluated as having a high risk of bias. Factors contributing to the risk of bias include inappropriate exclusion of participants (10/23, 43%), small sample size (15/23, 65%), poor handling of missing data (20/23, 87%), and incorrect method of selecting predictors (8/23, 35%). **Conclusions:** Most studies on nonlinear-machine-learning-based warfarin prediction models show poor methodological quality and have a high risk of bias. The analysis domain is the major contributor to the overall high risk of bias. External validity and model reproducibility are lacking in most studies. Future studies should focus on external validity, diminish risk of bias, and enhance real-world clinical relevance.



Citation: Zhang, F.; Liu, Y.; Ma, W.; Zhao, S.; Chen, J.; Gu, Z. Nonlinear Machine Learning in Warfarin Dose Prediction: Insights from Contemporary Modelling Studies. *J. Pers. Med.* **2022**, *12*, 717. <https://doi.org/10.3390/jpm12050717>

Academic Editors: Prahlad Ho and Amir Houshang Shemirani

Received: 14 March 2022

Accepted: 28 April 2022

Published: 29 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: warfarin; nonlinear machine learning; algorithms; model prediction; PROBAST

1. Introduction

Warfarin is an oral anticoagulant that is widely used in clinical practice for the treatment of venous thromboembolism (VTE) and thromboembolic events associated with atrial fibrillation (AF) or heart valve replacement (HVR) [1,2]. Despite its high effectiveness and low price, its optimal use is limited by a narrow therapeutic window and the highly

variable clinical response between individuals. Predicting optimal warfarin dose is influenced by many factors including demographic, clinical, environmental, and genetic factors, and multiple drug interactions, such as age, weight, height, vitamin K epoxide reductase complex subunit 1 (VKORC1) genotype, cytochrome P450 family 2 subfamily C member 9 (CYP2C9) genotype, coadministration of antiplatelet drugs, antimicrobials, nonsteroidal anti-inflammatory drugs (NSAIDs), and proton pump inhibitors, etc. [3–7]. Therefore, a tailored warfarin dose may help clinicians to select the appropriate dose for individualized warfarin treatment.

Many pharmacogenetic algorithms integrating clinical, demographic, and genetic variables, including linear regression (LR) models, physiologically based pharmacokinetic (PBPK) models, and nonlinear machine learning (ML) models, have been developed to predict the dose requirements of individual patients. The PBPK model can help decision making in relation to dose selection and clinical study strategies by identifying pharmacokinetic (PK) liabilities, such as poor bioavailability, high clearance, potential for drug–drug interactions (DDIs), or the need for dose adjustments in special populations [8–11]. However, due to the lack of *in vivo* data in patient populations, the distribution and absorption parameters could not be validated, which brought uncertainty to model parameters and outputs [8]. This hindered the practical application of the PBPK model. LR, a type of ML algorithm [12], is the most commonly used method because of its high interpretability [13]. However, the complex and nonlinear relationship between the factors mentioned above, warfarin responses, and metabolism render LR possible unsuitable to appropriately predict the warfarin maintenance dose [14,15]. Therefore, ML of nonlinear relationships between variables, warfarin response, and metabolism has been recently applied to enhance the model expression of the complicated relationship between the individual factors and warfarin dose.

Numerous published algorithm studies on warfarin dose prediction provide a plethora of information, and Asimwe et al. have examined the methodology and the risk of bias of warfarin dose prediction algorithms as of 20 May 2020 [13]. However, this article evaluated all dose prediction algorithms and did not focus on nonlinear ML algorithms, and new evidence may exist as of 2022.

Our systematic review seeks to provide a contemporary overview of the dosing algorithms for nonlinear ML research for clinical applications. We aimed to describe the study characteristics, evaluate the methods and quality of nonlinear ML studies, discuss the performance of nonlinear ML and LR algorithms, and account for the methodological considerations of nonlinear ML.

2. Methods

2.1. Search Strategy and Selection Criteria

This manuscript has been prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the corresponding checklist (Table S1). We performed a comprehensive search using medical subject headings and text words related to “warfarin” and “algorithm” to identify eligible studies (Table S2). Several electronic databases were searched from the inception to December 2021: PubMed, Embase, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI), China Biology Medicine (CBM), China Science and Technology Journal Database (VIP), and Wanfang Database. Additional articles were retrieved by manually searching the reference lists of relevant publications. A second search was conducted in March 2022 to identify the records published after our first search.

We selected the publications for review if they fulfilled the following inclusion criteria: (I) English or Chinese language; (II) observational or interventional studies that developed, validated, or assessed the warfarin dosing algorithm modelling with at least two predictor variables in any warfarin-treated population. The exclusion criteria included (I) LR algorithm studies; (II) physiologically based pharmacokinetic (PBPK) models for predicting the absorption, distribution, metabolism, and excretion (ADME) of synthetic or natural

chemical substances in humans; (III) informal publication types (such as commentaries, letters to the editor, editorials, and meeting abstracts). Warfarin dosing algorithms were defined as computational models that are composed of predictor variables to predict the weekly or daily dose of warfarin, including dose equations, nomograms, graphs, tables, and computer programs, etc.

2.2. Study Selection and Extraction of Data

After excluding the records that were irrelevant to our study, two reviewers (F.Z. and Y.L.) independently screened the abstracts for potentially eligible studies. The reports were then assessed for eligibility, and any disagreements were resolved by consensus. We designed a data extraction form based on four domains: participants, predictors, model development, and model evaluation. These were adapted from the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) [16] and the Prediction model Risk Of Bias Assessment Tool (PROBAST) [17]. Two reviewers (F.Z. and Y.L.) extracted the data from the study reports independently and in duplicate for each eligible study, and any disagreements were resolved by consensus or a third reviewer (Z.G.). When a study reported multiple nonlinear ML algorithms, each algorithm was extracted separately. When a study reported the same nonlinear machine learning algorithm based on different variables, we extracted the best model evaluation.

2.3. Risk of Bias

We assessed the risk of bias in each included study by applying the PROBAST. Although the above tools focus on prediction models that consider binary or time-to-event outcomes, the author encourages the use of these tools for other outcomes and other ML techniques. Additionally, many systematic reviews of ML algorithms also used PROBAST for risk of bias assessment [18–21]. PROBAST contains 20 signaling questions from four domains (participants, predictors, outcomes, and analyses), but we considered some problems to be less relevant to nonlinear ML studies (e.g., points that assigned weights in the final model). Generally, the algorithm does not receive information regarding the assigned weights. Therefore, we did not assess the questions regarding signaling (Table S3). Two reviewers (F.Z. and Y.L.) independently assessed the signaling questions according to the degree of compliance with the PROBAST recommendations. The disagreements were discussed until a consensus was reached. The risk-of-bias judgment for each domain was based on the answers to the signaling questions.

2.4. Data Synthesis

We did not conduct the formal quantitative syntheses because of the probable heterogeneity of specialties and outcomes. All extracted data were summarized and presented descriptively. All analyses were performed by Excel (version 2020).

3. Results

3.1. Study Selection

We aimed to provide a contemporary overview of dosing algorithms used in nonlinear ML research for warfarin. Figure 1 illustrates the literature search and selection process. Our electronic search retrieved 8996 records. Of these, we excluded 6134 records based on titles and abstracts. Of the 298 full-text records assessed for eligibility, 275 were excluded, and 23 studies were related to nonlinear ML.

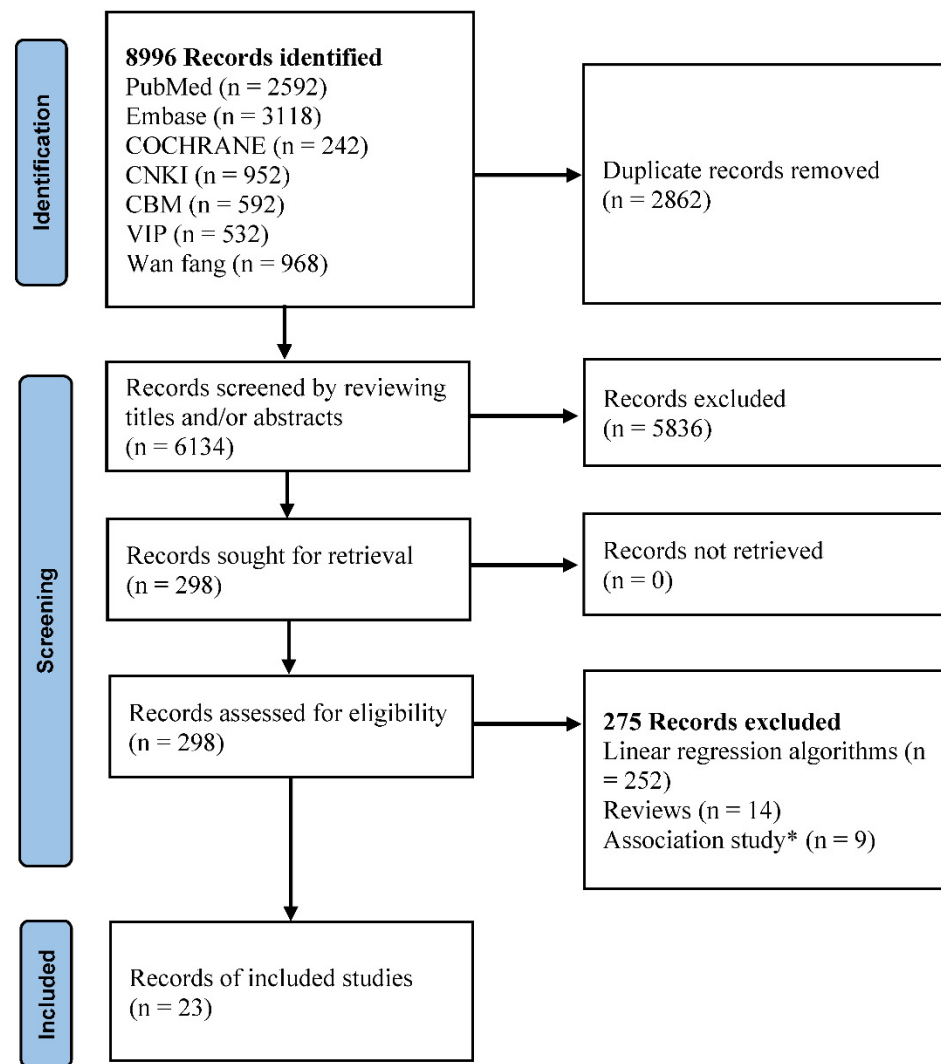


Figure 1. PRISMA flow chart of included studies. * The same research was translated into Chinese and English.

3.2. General Characteristics

Table 1 summarizes the general characteristics of 23 studies. Of these, 22 studies were retrospective.

Table 1. Summary characteristics of studies.

Study	Year	Study Type	Participants				Predictors			Model Development		Model Evaluation
			Source	Patients	Indication	Target INR *	Features	Features Selection	Missing Data Handling	Model Type	Machine Learning Algorithms #	Performance Measures
Solomon [22]	2004	Retrospective	Israel	148	NA	NA	clinical	Univariate analysis	NA	Development	NNM	r
Cosgun [23]	2011	Retrospective	USA	290	NA	2.0–3.0	clinical + genetic	Univariate analysis	single imputation	Development	DT, SV, Ensemble learning	R ²
Hu [24]	2012	Retrospective	China	587	NA	1.0–3.0	clinical	Expert opinion and literature review	NA	Development	DT, SV, KNN, Ensemble learning	MAE
Grossi [25]	2014	Retrospective	Italy	377	PE, DVT, AF, AHV, CM, Stroke, Others	2.0–4.0	clinical + genetic	Machine learning algorithm (TWIST system)	NA	Development	NNM	R ² , MAE, ideal dose
Saleh [26]	2014	Retrospective	IWPC sites	4271	PE, DVT, AF, AHV, CM, Stroke, Others	2.0–3.0	clinical + genetic	Backward Variable Selection	Excluded	Development	NNM	R ² , MAE, ideal dose
Zhou [27]	2014	Retrospective	China	1093	HVR	1.5–2.5	clinical	Univariate analysis, Stepwise regression	NA	Development	NNM	MAE, ideal dose
Li [28]	2015	Retrospective	China; IWPC sites	1511	HVR	1.7–3.0; 2.0–3.0	clinical + genetic	Stepwise regression	Excluded	External validation	DT, SV, NNM, Ensemble learning, Other	MAE, ideal dose
Liu [29]	2015	Retrospective	IWPC sites	4797	PE, DVT, AF, AHV, CM, Stroke, Others	2.0–3.0	clinical + genetic	Stepwise regression	Excluded	Development	DT, SV, NNM, Ensemble learning, Other	MAE, ideal dose
Alzubiedi [30]	2016	Retrospective	IWPC sites	163	PE, DVT, AF, Stroke, Others	2.0–3.0	clinical + genetic	Backward Variable Selection	NA	Development	NNM	R ² , MAE, ideal dose
Pavani [31]	2016	NR	India	240	PE, AF, HVR	No limitation	clinical + genetic	NA	NA	Development	NNM	R ² , MAE
Li [32]	2018	Retrospective	China	15,694	HVR	1.5–2.5	clinical	Covariance analysis, expert opinion, and literature review	NA	Development with external validation (same data)	NNM	MAE, RMSE, ideal dose

Table 1. Cont.

Study	Year	Study Type	Participants				Predictors			Model Development		Model Evaluation
			Source	Patients	Indication	Target INR *	Features	Features Selection	Missing Data Handling	Model Type	Machine Learning Algorithms #	Performance Measures
Ma [33]	2018	Retrospective	IWPC sites	5743	PE, DVT, AF, AHV, CM, Stroke, Others	1.7–3.3	clinical + genetic	Expert opinion and literature review	single imputation	Development	SV, NNM, Ensemble learning, KNN	MAE, ideal dose
Tao [34]	2018	Retrospective	China	13,639	HVR	1.5–2.5	clinical	Univariate analysis	NA	Development with external validation (same data)	NNM	MAE, MSE, ideal dose
Li [35]	2019	Retrospective	China	13,639	HVR	1.5–2.5	clinical	Univariate analysis	NA	Development with external validation (same data)	NNM	MAE, MSE, RMSE, ideal dose
Tao [36]	2019	Retrospective	China	289	NR	2.0–3.0	clinical + genetic	NA	NA	Development	NNM, SV, GP, Ensemble learning	R ² , MAE, MSE, ideal dose
Tao [37]	2019	Retrospective	China; IWPC sites	617	PE, DVT, AF, VR, ICT, EVE, stroke	2.0–3.0; 2.0–2.5	clinical + genetic	Expert opinion and literature review	NA	Development	DT, SV, Ensemble learning	R ² , MAE, MSE, ideal dose
Roche-Lima [38]	2020	Retrospective	USA	190	PE, DVT, AF, VR, DM2, CHF, Stroke, Others	2.0–3.0	clinical + genetic	NA	Excluded	Development	DT, SV, NNM, KNN, Ensemble learning, Other	MAE, ideal dose
Asimwe [39]	2021	Retrospective	Uganda, South Africa	634	AF, VT, VHT	2.5–3.5; 2.0–3.0	clinical	Expert opinion and literature review	multivariate imputation	Development with external validation (another data)	DT, SV, KNN, NNM, Ensemble learning, Other	MAE, MAPE, ideal dose
Gu [40]	2021	Retrospective	China	15,108	HVR	1.5–2.5	clinical	Univariate analysis	Excluded	Development with external validation (same data)	NNM	MAE, MSE, ideal dose

Table 1. Cont.

Study	Year	Study Type	Participants				Predictors			Model Development		Model Evaluation
			Source	Patients	Indication	Target INR *	Features	Features Selection	Missing Data Handling	Model Type	Machine Learning Algorithms #	Performance Measures
Liu [41]	2021	Retrospective	China	377	PE, DVT, AF, HF, PAH, Stroke	1.5–3.0	clinical + genetic	Univariate analysis	Not imputed	Development	Ensemble learning	R ² , MAE, MSE, RMSE, ideal dose
Ma [42]	2021	Retrospective	China	19,060	HVR	1.5–2.5	clinical	Univariate analysis	NA	Development with external validation (same data)	NNM	MAE, MSE, ideal dose
Nguyen [43]	2021	Retrospective	Korean	650	PE, DVT, HVR, VHD, Stroke, Arrhythmia, others	1.5–3.0	clinical + genetic	Recursive feature elimination	Single imputation	Development	Ensemble learning	r, MAE, RMSE, ideal dose
Steiner [44]	2021	Retrospective	IWPC sites, North and South America	7030	PE, DVT, TIA, Others	2.0–3.0; No limitation	clinical + genetic	NA	multivariate imputation	Development	DT, SV, Other	MAE, ideal dose

AF—atrial fibrillation; AHV—artificial heart valves; PE—pulmonary embolism; DVT—deep vein thrombosis; CM—cardiomyopathy; ICT—intracardiac thrombus; EVE—endovascular exclusion of aortic dissection; DM2—type 2 diabetes mellitus; CHF—congestive heart failure; HF—heart failure; VT—venous thromboembolism; VHT—valvular heart disease; PAH—pulmonary arterial hypertension; TIA—transient attack; HVR—heart valve replacement; INR—the international normalized ratio; NNM—neural network model; DT—decision tree; SV—support vector; KNN—K-nearest neighbor; GP—genetic programming; Other—other nonlinear regression model; r—coefficient of correlation; R²—coefficient of determination; MAE—mean absolute error; MSE—mean square error; RMSE—root mean square error. * In the article, there were different target INRs based on different indications in the same dataset, we took the minimum and maximum value of the target INRs. There were different target INRs in different datasets, we took the target INRs separately. # The nonlinear machine learning algorithms were divided into 6 categories (DT, SV, NNM, KNN, GP, ensemble learning, other nonlinear regression) based on the algorithms involved in the studies. When a study reported many subcategories in a large category, the large category was reported in the table.

3.2.1. Participants

The Asian and the International Warfarin Pharmacogenetics Consortium (IWPC) sites accounted for 78% of studies, with the top two countries being China (9/23, 39%) and the USA (2/23, 9%). The median sample size for these studies was 650 (range 148–19,060). A total of 19 (83%) studies clarified the indications. Of 23 studies, 20 (87%) studies described the target international normalized ratio (INR), which is a parameter that quantifies the coagulation activity using the prothrombin time and is used for the regular monitoring of warfarin therapy.

3.2.2. Predictors

All studies explored the demographic and clinical predictors, and 14 of the 23 (61%) studies explored the genetic factors. The predictors that included at least four factors are shown in Figure 2. Age (21/23, 91%), weight (17/23, 74%), height (12/23, 52%), and amiodarone combination (11/23, 48%) were the four most common demographic and clinical predictors. CYP2C9 (14/23, 61%), VKORC1(14/23, 61%), and CYP4F2 (5/23, 22%) were the three most common genetic predictors. With respect to features selection, the top three methods were as follows: univariate analysis (8/23, 35%), expert opinion and literature review (5/23, 22%), and stepwise regression (3/23, 22%). For missing data handling, a majority of the studies (17/23, 74%) did not report or deleted directly.

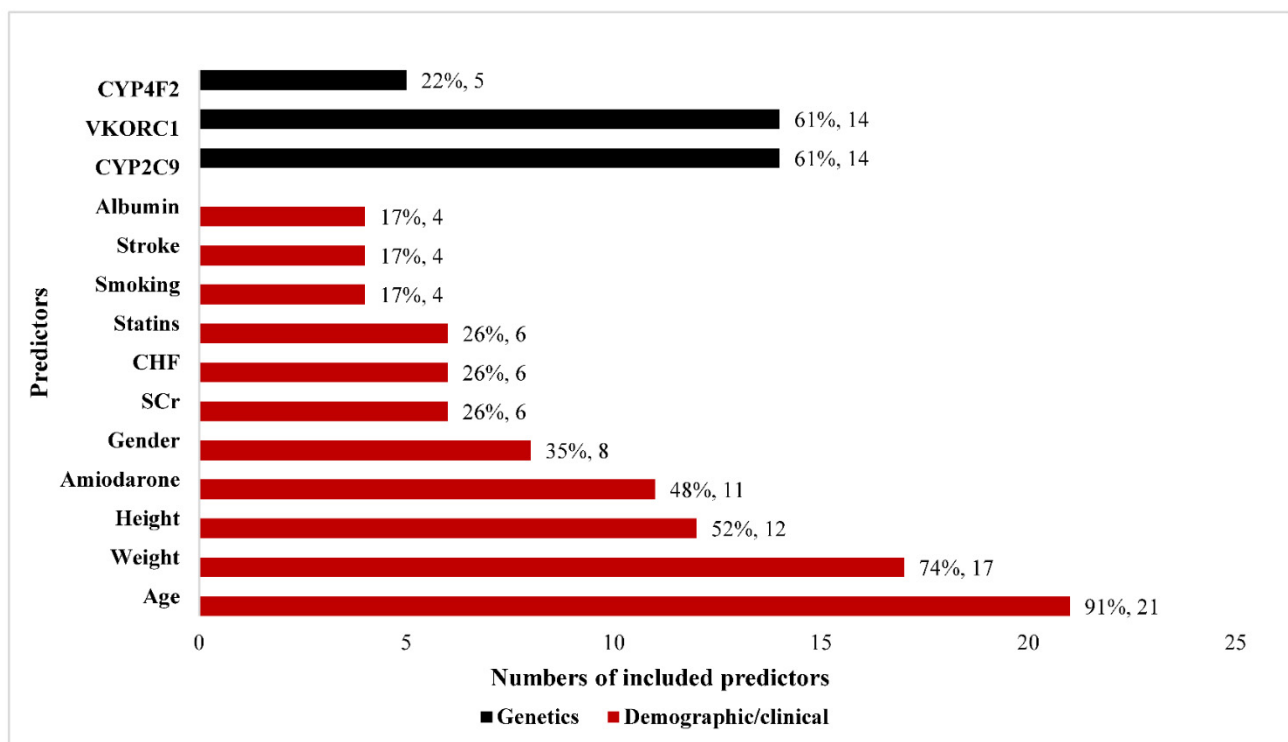


Figure 2. Predictors included at least 4 factors in all studies. SCr—serum creatinine; CHF—congestive heart failure.

3.2.3. Model Development

The prediction model studies included were categorized into four types: development (with internal validation), development with external validation (same data), development with external validation (another data), and external validation only. Regarding the model type, 16 of the 23 (70%) studies were development studies, 5 of the 23 (22%) studies were development with external validation (same data) studies, 1 was a development with external validation (another data) study, and 1 was an external validation study. The single neural network model (NNM) algorithm was used for model development in 6/11 (55%)

studies and was used for development with external validation (same data) in 5/11 (45%) studies. Algorithms (2 or more) (e.g., DT (decision tree), SVR (support vector regression), KNN (K-nearest neighbor), ensemble learning, and other nonlinear regression model) were used for development in 10/12 (83%) studies; 1 was used for development with external validation (another data) and 1 was used for external validation.

3.2.4. Model Evaluation

Of the 23 studies, 8 (35%) studies described the R^2 (the coefficient of determination), which is a parameter fit for accuracy measurement that represents the proportion of total interpatient variability in warfarin dose requirements, accounted for by the variables included in the algorithm. Regarding the precision (predictive accuracy), the most reported measures were MAE (the mean absolute error) in 21/23 (91%) and RMSE (the root mean square error) or the mean square error (MSE) were used in 9/23 (39%). Of the 23 studies, 19 (83%) studies described the ideal dose (the absolute prediction error between predicted dose and the actual dose was within 20% of the actual dose).

Of the included studies, the MAE ranged from 1.47 to 10.86 mg/week in development studies, from 2.42 to 5.18 mg/week in development with external validation (same data), from 12.07 to 17.59 mg/week in development with external validation (another data), and from 4.40 to 4.84 mg/week in external validation (Table 2).

Table 2. Performance evaluation of the included nonlinear machine learning algorithms studies.

Studies	NO. Models	Models	NO. Patients	NO. Features	MAE (mg/Week)
Development					
Solomon 2004	1	NNM	148	3	NR
Cosgun 2011	3	DT, SV, Ensemble learning	290	11	NR
Hu 2012	9	DT, SV, KNN, Ensemble learning	587	7	(1.47, 1.55)
Grossi 2014	1	NNM	377	14	5.72
Saleh 2014	1	NNM	4271	9	9
Zhou 2014	1	NNM	1093	11	0.08 *
Liu 2015	7	DT, SV, NNM, Ensemble learning, Other	4797	9	(8.84, 9.82)
Alzubiedi 2016	1	NNM	163	7	11.2
Pavani 2016	1	NNM	240	9	−1.97 *
Ma 2018	8	SV, NNM, KNN, Ensemble learning	5743	13	(8.31, 10.86)
Tao 2019	6	NNM, SV, GP, Ensemble learning	289	7	NR
Tao 2019	4	DT, SV, Ensemble learning	617	11	(4.73, 5.36)
Roche-Lima 2020	9	DT, SV, NNM, KNN, Ensemble learning, Other	190	24	(4.73, 9.87)
Liu 2021	3	Ensemble learning	377	11	(2.98, 4.54)
Nguyen 2021	1	Ensemble learning	650	17	4.48
Steiner 2021	3	DT, SV, Other	7030	13	(8.11, 8.18)
Development with external validation (same data)					
Li 2018 IV	1	NNM	15,694	12	2.59
Li 2018 EV	1	NNM	15,694	12	2.68
Tao 2018 IV	1	NNM	13,639	9	4.07
Tao 2018 EV	1	NNM	13,639	9	4.22
Li 2019 IV	1	NNM	13,639	10	4.82
Li 2019 EV	1	NNM	13,639	10	5.18
Gu 2021 IV	1	NNM	15,108	8	2.58
Gu 2021 EV	1	NNM	15,108	8	2.59
Ma 2021 IV	2	NNM	19,060	8	(2.28, 3.04)
Ma 2021 EV	2	NNM	19,060	8	(2.42, 2.88)

Table 2. Cont.

Studies	NO. Models	Models	NO. Patients	NO. Features	MAE (mg/Week)
Development with external validation (another data)					
Asimwe 2021	13	DT, SV, KNN, NNM, Ensemble learning, Other	270	7	(12.07, 17.59)
External validation					
Li 2015	6	DT, SV, NNM, Ensemble learning, Other	1295	10	(4.41, 4.76)
Li 2015	6	DT, SV, NNM, Ensemble learning, Other	216	10	(4.40, 4.84)

NNM—neural network model; DT—decision tree; SV—support vector; KNN—K-nearest neighbor; GP—genetic programming. * The value provided in literature was not clear, and it was impossible to distinguish whether it was derived from the training set or the test set.

3.3. Methods and Risk of Bias

The overall and detailed risk of bias are presented in Figure 3 and Table S4. Specifically, 8 (35%) studies were rated to have a high risk of bias in the participant domain. The major issue in this domain is the inappropriate exclusion of participants (43%); 1 (4%) study was considered as having high risk of bias in the predictor domain; 8 (35%) studies had high risk of bias in the outcome domain; and 22 (96%) studies were judged as high risk of bias in the analysis domain (Table 3).

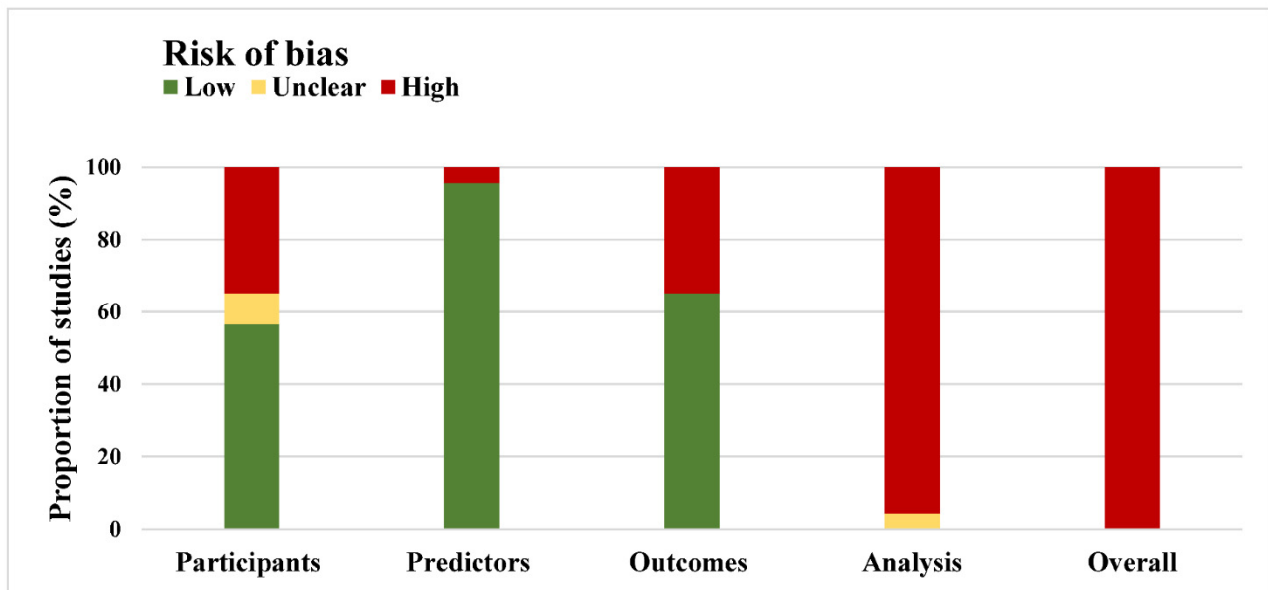


Figure 3. PROBAST (Prediction Model Risk of Bias Assessment Tool) risk of bias assessment for 23 studies.

Table 3. PROBAST signaling questions in 23 included studies.

Signaling Question No.	Signaling Question	Included Studies (n = 23)		
		Yes or Probably Yes	No or Probably No	No Information
Participant domain		number (percentage, 95% confidence interval)		
1.1	Were appropriate data sources used, e.g., cohort, RCT, or nested case-control study data?	23 (100, 100 to 100)	0	0
1.2	Were all inclusions and exclusions of participants appropriate?	13 (57, 36 to 77)	8 (35, 15 to 54)	2 (8, 3 to 20)
Predictor domain				
2.1	Were predictors defined and assessed in a similar way for all participants?	23 (100, 100 to 100)	0	0
2.2	Were predictor assessments made without knowledge of outcome data?	22 (96, 87 to 100)	1 (4, 4 to 13)	0
2.3	Are all predictors available at the time the model is intended to be used?	22 (96, 87 to 100)	1 (4, 4 to 13)	0
Outcome domain				
3.1	Was the outcome determined appropriately?	16 (70, 51 to 89)	7 (30, 12 to 49)	0
3.3	Were predictors excluded from the outcome definition?	22 (96, 87 to 100)	1 (4, 4 to 13)	0
3.4	Was the outcome defined and determined in a similar way for all participants?	21 (91, 80 to 100)	0	2 (9, 3 to 20)
3.5	Was the outcome determined without knowledge of predictor information?	22 (96, 87 to 100)	1 (4, 4 to 13)	0
3.6	Was the time interval between predictor assessment and outcome determination?	23 (100, 100 to 100)	0	0
Analysis domain				
4.1	Were there a reasonable number of participants with the outcome?	8 (35, 15 to 54)	12 (52, 32 to 73)	3 (13, 1 to 27)
4.3	Were all enrolled participants included in the analysis?	23 (100, 100 to 100)	0	0
4.4	Were participants with missing data handled appropriately?	3 (13, 1 to 27)	20 (87, 73 to 100)	0
4.5	Was selection of predictors based on univariable analysis avoided?	11 (48, 27 to 68)	8 (35, 15 to 54)	4 (17, 2 to 33)
4.7	Were relevant model performance measures evaluated appropriately?	21 (91, 80 to 100)	2 (9, 3 to 20)	0
4.8	Were model overfitting and optimism in model performance accounted for?	19 (90, 78 to 100)	2 (10, 3 to 22)	0

Signaling questions 3.2, 4.2, 4.6, and 4.9 were not included (Table S3). The risk of bias judgment for each domain was based on the answers to the signaling questions. If the answer to all signaling questions was yes or probably yes, then the domain was judged as low risk of bias. If reported information was insufficient to answer the signaling questions, these were judged as no information. If more than half of the answer to all signaling questions were judged as no information, then the domain was judged as high risk of bias, otherwise the domain was judged as unclear risk of bias. If one answer to all signaling question was answered as no or probably no, then the domain was judged as high risk of bias. After judging all the domains, we performed an overall assessment for each application of PROBAST. This tool recommends rating the study as low risk of bias if all domains had low risk of bias. If at least one domain had a high risk of bias, overall judgment was rated as high risk of bias. If the risk of bias was unclear in at least one domain and all other domains had a low risk of bias, then an unclear risk of bias was assigned.

Thus, the analysis domain was major contributor to the overall high risk of bias. In detail, the number of participants with the outcome was unreasonable or unclear in 15 (65%) studies; 20 (87%) studies were inappropriate regarding the handling of missing data (missing data were omitted or excluded); 8 (35%) studies used univariable analyses to

select the predictors. Given above, the factors contributing to the risk of bias include the inappropriate exclusion of participants (10/23, 43%), small sample size (15/23, 65%), poor handling of missing data (20/23, 87%), and incorrect method of selecting predictors (8/23, 35%) (Table 3).

4. Discussion

4.1. Major Finding

In this study, we summarized the warfarin dosing algorithms for nonlinear ML based on their general characteristics and risk of bias. The major findings were as follows: (I) Most studies on nonlinear-machine-learning-based warfarin prediction models show poor methodological quality and are at high risk of bias, mainly as a result of inappropriate exclusions, small sample size, inappropriate handling of participants with missing data, and incorrect predictor selection; (II) About 50% of the involved studies focused on the Asian population, and the included clinical variables (age, weight, height, amiodarone combination) and genetic variables (CYP2C9, VKORC1) were recognized by most studies; (III) Most studies were limited to model internal validation, and external validation was rarely used to further illustrate the generalization ability of the predictive model.

4.2. Risk of Bias

Overall, most studies show poor methodological quality and are at a high risk of bias. Without a rigorous evidence base, these research results may not be applicable to the clinical use and should rather guide their efforts towards improving the design and quality [45]. The following key improvements were established from our methodological review.

First, inappropriately excluding participants may result in the target population not being represented. The exclusion criteria of 7/23 (30%) studies included complications that occurred during anticoagulant therapy (thrombosis, embolism, bleeding, death result) or severe liver and kidney dysfunction after the operation, which may introduce a bias because the final study population represented a selected, lower-risk sample of the original population. This may produce biased estimates of the predictive performance of the model. Therefore, reasonable inclusion and exclusion criteria should be adopted.

Second, since EPV (the number of events per candidate predictor variable) may not be the best method for nonlinear ML in warfarin dose prediction, we applied more stringent criteria for signal identification (EPV > 200 [46]). We also found that the training set size of studies with EPV rated “N” was small, ranging from 108 to 587. This size is small for ML, making it impossible for us to judge the potential biases. Although no studies have been conducted regarding the sample size calculations for developing prediction models using nonlinear ML techniques, these studies usually require more participants than conventional statistical approaches do [47].

Third, with respect to missing data, most studies deleted or failed to provide the methods. Multiple imputation is generally preferred because it prevents biased model performance as a result of deletion or single imputation of participants’ missing data. However, in warfarin prediction models with nonlinear ML techniques, multiple imputations remain a minority. Therefore, it would be useful if algorithm developers could improve the imputation methods in their models when possible.

Forth, univariate analysis to select the predictors may be an incorrect method of predictor selection because predictors are chosen on the basis of their statistical significance as a single predictor rather than in context with other predictors [48]. Well-established predictors should be included and retained in the model, regardless of any statistical significance. Therefore, we recommend that researchers avoid the use of univariable analysis to select the candidate predictors and choose nonstatistical methods (without any statistical univariable pretesting of the associations between candidate predictors and outcome) or other methods (for example, principal component analysis and lasso regression).

4.3. Performance Measures

We originally aimed to evaluate all the measurement parameters of the algorithms. As reported previously, the coefficient of determination (R^2) was the most common parameter fit for accuracy measurement, and MAE, MSE, and RMSE were the most common parameters for predictive accuracy measurement. However, as a measure of the degree of association between actual and predicted doses, R^2 may be misleading in terms of the actual closeness of predictions to the true values. Hence, it is not a good measure of the predictive accuracy [49,50]. As the proximity distance measures between the actual and predicted doses, accuracy measure parameters (MAE, MSE, RMSE) have the following issue: 1 mg/d may be clinically more important in a small value than a similar error in a large value. For example, numerous studies have shown that the Asian populations display higher sensitivity to the anticoagulant effects of warfarin, and the overall stable dose is relatively small compared with Caucasian populations. The MAE made by the model is correspondingly smaller than that of the other races. However, this does not mean that the model is better. These values should be interpreted cautiously.

4.4. Linear Regression vs. Nonlinear ML in Warfarin

As a traditional and widely used oral anticoagulant, warfarin dose prediction models have been studied for a long time [13]. LR was originally used to develop a warfarin predictive model because of its simple development process and high interpretability. Gage et al. and the International Warfarin Pharmacogenetics Consortium (IWPC) have developed two representative linear warfarin prediction models based on pharmacogenomic information and clinical factors [4,51]. However, the LR method presents certain irreconcilable issues, such as poor behaviour of the nonlinear relationship between variables, which makes the LR model an inappropriate method [14,15]. With the development of science and technology, nonlinear ML has gradually emerged as a powerful technique for analysing complex analytic problems and could use nonlinear, highly interactive combinations of predictors to uncover novel patterns that may improve the predictive performance [52–54]. Therefore, an increasing number of researchers are beginning to develop warfarin predictive models using nonlinear ML methods to explore new possibilities.

However, identifying which algorithm performs better (either LR or nonlinear ML) is difficult, because these algorithms are derived from different racial backgrounds. Therefore, we compared nonlinear ML with LR in the same population and variables (Table S5). Overall, the results of these studies still cannot determine which algorithm performs better. In some studies, the ML had a lower MAE than the LR, and vice versa. With these points in mind, further development and validation of warfarin dose prediction models based on nonlinear ML, and a more thorough comparison with LR, are recommended.

4.5. Clinical Relevance

In the included literature, we also extracted the suggestion in clinical practice (suggestion in discussion that algorithm can now be used clinically). Unfortunately, few studies underwent the description of the practical application using the developed model. Therefore, a simple guide process aiming at the population of anticoagulant therapy using warfarin might be necessary. First, surgeons should conduct a comprehensive assessment of patient conditions for the risk factors before and after procedure. Patients who are older, more fragile, and have the VKORC1 or CYP2C9 gene carrier might be at high risk of bleeding [55]. In addition, concomitant drugs to warfarin therapy may also increase the risk of bleeding, such as amiodarone, coadministration of antiplatelet drugs, antimicrobials, etc. [56]. Second, when determining warfarin doses during treatment, we suggest using externally validated dosing algorithms over an internal validated dosing algorithms or an ad hoc approach. Certainly, the dosing algorithms that have been externally validated across races, which focuses on a certain race and extends to different races via appropriate weighting or modifications, are more suitable for clinical application. Third, patients with extremely high or low doses determined by the algorithm should receive more frequent

INR testing and bleeding risk assessment after surgery. Forth, we suggest enrolling patients with warfarin in a structured care process, which may include sophisticated patient tracking systems, comprehensive patient education, outcome evaluation, and quality improvement activities.

4.6. Study Limitations

The main limitations of this study are as follows: first, the nomenclature in the field is sometimes used in a non-standardized manner, and thus some potentially eligible studies might have included terminologies that were not captured by our search strategy. Although comprehensive, our search might have missed some studies that could have been included. Second, the benchmark that we adopted to evaluate the risk of bias (PROBAST) was designed for conventional prediction modelling studies, this has certain limitations. Third, the risk of bias entails some subjective judgment, and people with different experiences may have varying perceptions.

4.7. Future Research

The number of warfarin dose prediction models based on nonlinear ML is increasing every year; thus, their identification, reporting, and assessment have become even more relevant. We believe that researchers should consider the following points: first, researchers should consider the impact of different populations, indications, and valve types on the range of INR treatment range, whether the inclusion and exclusion criteria of the included participants can accurately represent the target population or not, appropriate handling of participants with missing data, and correct predictor selection methods. Second, with respect to model development, we recommend that the researchers report detailed information about modelling methods and encourage code sharing to enhance the possibility of reproducibility. Third, for the better judgement of studies, we recommend that researchers should adhere to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement [57]. Although TRIPOD was not explicitly developed for machine learning prediction models, all items were applicable.

5. Conclusions

Most warfarin dosing algorithms for nonlinear ML have been developed in Asian populations, and few algorithms have been externally validated. Most studies show poor methodological quality and are at a high risk of bias, which makes them unreliable for clinical use. Factors contributing to the risk of bias include the inappropriate exclusion of participants, small sample size, poor handling of missing data, and incorrect method of selecting predictors. Efforts to improve the design, conduct, reporting, and validation studies of warfarin dosing algorithms of nonlinear ML are necessary to boost its application in clinical practice.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm12050717/s1>. Table S1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement, Table S2: Search strategy used in PubMed, Embase, and Cochrane, Table S3: PROBAST altered or excluded items, Table S4: Risk of Bias table: based on authors' judgements about each risk of bias item, Table S5. Comparison of Non-linear machine learning algorithms and Linear regression algorithms in the same population for Warfarin dose prediction.

Author Contributions: J.C., and Z.G. wrote and reviewed the entire manuscript; F.Z. and Y.L. contributed to the study conception and design, and critically reviewed the manuscript for important intellectual content; W.M. and S.Z. contributed to the data acquisition, analysis, and interpretation. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (71974137), Research Project of Drug Clinical Comprehensive Evaluation and Drug Treatment Pathway (SHYXH-ZP-2021-001), Clinical Research Innovation and Cultivation Fund of Ren Ji Hospital (RJPY-LX-008),

Ren Ji Boost Project of National Natural Science Foundation of China (RJTJ-JX-001), and Shanghai “Rising Stars of Medical Talent” Youth Development Program—Youth Medical Talents, Clinical Pharmacist Program (SHWJRS (2019)_072).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on reasonable request from the corresponding authors.

Conflicts of Interest: No conflicts of interest to declare.

References

- Nishimura, R.A.; Otto, C.M.; Bonow, R.O.; Carabello, B.A.; Erwin, J.P.; Fleisher, L.A.; Jneid, H.; Mack, M.J.; McLeod, C.J.; O’Gara, P.T.; et al. 2017 AHA/ACC Focused Update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* **2017**, *135*, e1159–e1195. [[CrossRef](#)] [[PubMed](#)]
- Witt, D.M.; Clark, N.P.; Kaatz, S.; Schnurr, T.; Ansell, J.E. Guidance for the practical management of warfarin therapy in the treatment of venous thromboembolism. *J. Thromb. Thrombolysis* **2016**, *41*, 187–205. [[CrossRef](#)] [[PubMed](#)]
- Loebstein, R.; Yonath, H.; Peleg, D.; Almog, S.; Rotenberg, M.; Lubetsky, A.; Roitelman, J.; Harats, D.; Halkin, H.; Ezra, D. Interindividual variability in sensitivity to warfarin—Nature or nurture? *Clin. Pharmacol. Ther.* **2001**, *70*, 159–164. [[CrossRef](#)] [[PubMed](#)]
- Klein, T.E.; Altman, R.B.; Eriksson, N.; Gage, B.F.; Kimmel, S.E.; Lee, M.T.M.; Limdi, N.A.; Page, D.; Roden, D.M.; Wagner, M.J.; et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* **2009**, *360*, 753–764. [[CrossRef](#)] [[PubMed](#)]
- Bourgeois, S.; Jorgensen, A.; Zhang, E.J.; Hanson, A.; Gillman, M.S.; Bumpstead, S.; Toh, C.H.; Williamson, P.; Daly, A.K.; Kamali, F.; et al. A multi-factorial analysis of response to warfarin in a UK prospective cohort. *Genome Med.* **2016**, *8*, 2. [[CrossRef](#)]
- Spina, E.; Barbieri, M.A.; Cicala, G.; Bruno, A.; de Leon, J. Clinically relevant drug interactions between newer antidepressants and oral anticoagulants. *Expert Opin. Drug Metab. Toxicol.* **2020**, *16*, 31–44. [[CrossRef](#)]
- Wang, M.; Zeraatkar, D.; Obeda, M.; Lee, M.; Garcia, C.; Nguyen, L.; Agarwal, A.; Al-Shalabi, F.; Benipal, H.; Ahmad, A.; et al. Drug-drug interactions with warfarin: A systematic review and meta-analysis. *Br. J. Clin. Pharmacol.* **2021**, *87*, 4051–4100. [[CrossRef](#)]
- Sager, J.E.; Yu, J.; Ragueneau-Majlessi, I.; Isoherranen, N. Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation Approaches: A Systematic Review of Published Models, Applications, and Model Verification. *Drug Metab. Dispos.* **2015**, *43*, 1823–1837. [[CrossRef](#)]
- Zhu, Y.-B.; Hong, X.-H.; Wei, M.; Hu, J.; Chen, X.; Wang, S.-K.; Zhu, J.-R.; Yu, F.; Sun, J.-G. Development of a novel individualized warfarin dose algorithm based on a population pharmacokinetic model with improved prediction accuracy for Chinese patients after heart valve replacement. *Acta Pharmacol. Sin.* **2017**, *38*, 434–442. [[CrossRef](#)]
- Arwood, M.J.; Deng, J.; Drozda, K.; Pugach, O.; Nutescu, E.A.; Schmidt, S.; Duarte, J.D.; Cavallari, L.H. Anticoagulation endpoints with clinical implementation of warfarin pharmacogenetic dosing in a real-world setting: A proposal for a new pharmacogenetic dosing approach. *Clin. Pharmacol. Ther.* **2017**, *101*, 675–683. [[CrossRef](#)]
- Lu, Y.; Yang, J.; Zhang, H.; Yang, J. Prediction of warfarin maintenance dose in Han Chinese patients using a mechanistic model based on genetic and non-genetic factors. *Clin. Pharmacokinet.* **2013**, *52*, 567–581. [[CrossRef](#)] [[PubMed](#)]
- DeGregory, K.W.; Kuiper, P.; DeSilvio, T.; Pleuss, J.D.; Miller, R.; Roginski, J.W.; Fisher, C.B.; Harness, D.; Viswanath, S.; Heymsfield, S.B.; et al. A review of machine learning in obesity. *Obes. Rev.* **2018**, *19*, 668–685. [[CrossRef](#)] [[PubMed](#)]
- Asiimwe, I.G.; Zhang, E.J.; Osanlou, R.; Jorgensen, A.L.; Pirmohamed, M. Warfarin dosing algorithms: A systematic review. *Br. J. Clin. Pharmacol.* **2021**, *87*, 1717–1729. [[CrossRef](#)] [[PubMed](#)]
- Schalekamp, T.; Brassé, B.; Roijers, J.; van Meegen, E.; van der Meer, F.; van Wijk, E.; Egberts, A.; de Boer, A. VKORC1 and CYP2C9 genotypes and phenprocoumon anticoagulation status: Interaction between both genotypes affects dose requirement. *Clin. Pharmacol. Ther.* **2007**, *81*, 185–193. [[CrossRef](#)] [[PubMed](#)]
- Li, X.; Liu, R.; Yan, H.; Tang, J.; Yin, J.; Mao, X.; Yang, F.; Luo, Z.; Tan, S.; He, H.; et al. Effect of CYP2C9-VKORC1 interaction on warfarin stable dosage and its predictive algorithm. *J. Clin. Pharmacol.* **2015**, *55*, 251–257. [[CrossRef](#)] [[PubMed](#)]
- Moons, K.; de Groot, J.; Bouwmeester, W.; Vergouwe, Y.; Mallett, S.; Altman, D.; Reitsma, J.; Collins, G. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med.* **2014**, *11*, e1001744. [[CrossRef](#)]
- Wolff, R.; Moons, K.; Riley, R.; Whiting, P.; Westwood, M.; Collins, G.; Reitsma, J.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [[CrossRef](#)]
- Andaur Navarro, C.L.; Damen, J.A.A.G.; Takada, T.; Nijman, S.W.J.; Dhiman, P.; Ma, J.; Collins, G.S.; Bajpai, R.; Riley, R.D.; Moons, K.G.; et al. Protocol for a systematic review on the methodological and reporting quality of prediction model studies using machine learning techniques. *BMJ Open* **2020**, *10*, e038832. [[CrossRef](#)]

19. Nagendran, M.; Chen, Y.; Lovejoy, C.A.; Gordon, A.C.; Komorowski, M.; Harvey, H.; Topol, E.J.; Ioannidis, J.P.A.; Collins, G.S.; Maruthappu, M. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* **2020**, *368*, m689. [[CrossRef](#)]
20. Kennedy, E.E.; Bowles, K.H.; Aryal, S. Systematic review of prediction models for postacute care destination decision-making. *J. Am. Med. Inform. Assoc.* **2021**, *29*, 176–186. [[CrossRef](#)]
21. van de Sande, D.; van Genderen, M.E.; Huiskens, J.; Gommers, D.; van Bommel, J. Moving from bytes to bedside: A systematic review on the use of artificial intelligence in the intensive care unit. *Intensiv. Care Med.* **2021**, *47*, 750–760. [[CrossRef](#)] [[PubMed](#)]
22. Solomon, I.; Maharshak, N.; Chechik, G.; Leibovici, L.; Lubetsky, A.; Halkin, H.; Ezra, D.; Ash, N. Applying an artificial neural network to warfarin maintenance dose prediction. *Isr. Med. Assoc. J.* **2004**, *6*, 732–735. [[PubMed](#)]
23. Cosgun, E.; Limdi, N.A.; Duarte, C.W. High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. *Bioinformatics* **2011**, *27*, 1384–1389. [[CrossRef](#)] [[PubMed](#)]
24. Hu, Y.H.; Wu, F.; Lo, C.L.; Tai, C.T. Predicting warfarin dosage from clinical data: A supervised learning approach. *Artif. Intell. Med.* **2012**, *56*, 27–34. [[CrossRef](#)]
25. Grossi, E.; Podda, G.M.; Pugliano, M.; Gabba, S.; Verri, A.; Carpani, G.; Buscema, M.; Casazza, G.; Cattaneo, M. Prediction of optimal warfarin maintenance dose using advanced artificial neural networks. *Pharmacogenomics* **2014**, *15*, 29–37. [[CrossRef](#)]
26. Saleh, M.I.; Alzubiedi, S. Dosage individualization of warfarin using artificial neural networks. *Mol. Diagn. Ther.* **2014**, *18*, 371–379. [[CrossRef](#)]
27. Zhou, Q.; Kwong, J.; Chen, J.; Qin, W.; Chen, J.; Dong, L. Use of artificial neural network to predict warfarin individualized dosage regime in Chinese patients receiving low-intensity anticoagulation after heart valve replacement. *Int. J. Cardiol.* **2014**, *176*, 1462–1464. [[CrossRef](#)]
28. Li, X.; Liu, R.; Luo, Z.Y.; Yan, H.; Huang, W.H.; Yin, J.Y.; Mao, X.Y.; Chen, X.P.; Liu, Z.Q.; Zhou, H.H.; et al. Comparison of the predictive abilities of pharmacogenetics-based warfarin dosing algorithms using seven mathematical models in Chinese patients. *Pharmacogenomics* **2015**, *16*, 583–590. [[CrossRef](#)]
29. Liu, R.; Li, X.; Zhang, W.; Zhou, H.H. Comparison of Nine Statistical Model Based Warfarin Pharmacogenetic Dosing Algorithms Using the Racially Diverse International Warfarin Pharmacogenetic Consortium Cohort Database. *PLoS ONE* **2015**, *10*, e0135784. [[CrossRef](#)]
30. Alzubiedi, S.; Saleh, M.I. Pharmacogenetic-guided Warfarin Dosing Algorithm in African-Americans. *J. Cardiovasc. Pharmacol.* **2016**, *67*, 86–92. [[CrossRef](#)]
31. Pavani, A.; Naushad, S.M.; Kumar, R.M.; Srinath, M.; Malempati, A.R.; Kutala, V.K. Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization. *Pharmacogenomics* **2016**, *17*, 121–131. [[CrossRef](#)] [[PubMed](#)]
32. Li, Q.; Tao, H.; Wang, J.; Zhou, Q.; Chen, J.; Qin, W.Z.; Dong, L.; Fu, B.; Hou, J.L.; Chen, J.; et al. Warfarin maintenance dose Prediction for Patients undergoing heart valve replacement- a hybrid model with genetic algorithm and Back-Propagation neural network. *Sci. Rep.* **2018**, *8*, 9712. [[CrossRef](#)] [[PubMed](#)]
33. Ma, Z.; Wang, P.; Gao, Z.; Wang, R.; Khalighi, K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLoS ONE* **2018**, *13*, e0205872. [[CrossRef](#)] [[PubMed](#)]
34. Tao, H.; Li, Q.; Zhou, Q.; Chen, J.; Fu, B.; Wang, J.; Qin, W.; Hou, J.; Chen, J. A prediction study of warfarin individual stable dose after mechanical heart valve replacement: Adaptive neural-fuzzy inference system prediction. *PLoS ONE* **2018**, *18*, 10. [[CrossRef](#)] [[PubMed](#)]
35. Li, Q.; Wang, J.; Tao, H.; Zhou, Q.; Chen, J.; Fu, B.; Qin, W.Z.; Li, D.; Hou, J.L.; Chen, J.; 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.* **2019**, *40*, 41–53. [[CrossRef](#)]
36. Tao, Y.; Chen, Y.J.; Fu, X.; Jiang, B.; Zhang, Y. Evolutionary Ensemble Learning Algorithm to Modeling of Warfarin Dose Prediction for Chinese. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 395–406. [[CrossRef](#)]
37. Tao, Y.; Chen, Y.J.; Xue, L.; Xie, C.; Jiang, B.; Zhang, Y. An Ensemble Model With Clustering Assumption for Warfarin Dose Prediction in Chinese Patients. *Pharm. J.* **2019**, *23*, 2642–2654. [[CrossRef](#)]
38. Roche-Lima, A.; Roman-Santiago, A.; Feliu-Maldonado, R.; Rodriguez-Maldonado, J.; Nieves-Rodriguez, B.G.; Carrasquillo-Carrion, K.; Ramos, C.M.; Da Luz Sant’Ana, I.; Massey, S.E.; Duconge, J. Machine learning algorithm for predicting warfarin dose in caribbean hispanics using pharmacogenetic data. *Front. Pharmacol.* **2020**, *10*, 1550. [[CrossRef](#)]
39. Asiimwe, I.G.; Blockman, M.; Cohen, K.; Cupido, C.; Hutchinson, C.; Jacobson, B.; Lamorde, M.; Morgan, J.; Mouton, J.P.; Nakagaayi, D.; 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 Pharmacomet. Syst. Pharmacol.* **2021**, *11*, 20–29. [[CrossRef](#)]
40. Gu, Z.C.; Huang, S.R.; Dong, L.; Zhou, Q.; Wang, J.; Fu, B.; Chen, J. 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. *Cardiovasc. Drugs Ther.* **2021**. [[CrossRef](#)]
41. Liu, Y.; Chen, J.; You, Y.; Xu, A.; Li, P.; Wang, Y.; Sun, J.; Yu, Z.; Gao, F.; Zhang, J. 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. [[CrossRef](#)] [[PubMed](#)]

42. Ma, W.; Li, H.; Dong, L.; Zhou, Q.; Fu, B.; Hou, J.-L.; Wang, J.; Qin, W.; Chen, J. Warfarin maintenance dose prediction for Chinese after heart valve replacement by a feedforward neural network with equal stratified sampling. *Sci. Rep.* **2021**, *11*, 13778. [[CrossRef](#)] [[PubMed](#)]
43. Nguyen, V.L.; Nguyen, H.D.; Cho, Y.S.; Kim, H.S.; Han, I.Y.; Kim, D.K.; Ahn, S. Comparison of multivariate linear regression and a machine learning algorithm developed for prediction of precision warfarin dosing in a Korean population. *J. Thromb. Haemost.* **2021**, *19*, 1676–1686. [[CrossRef](#)] [[PubMed](#)]
44. Steiner, H.E.; Giles, J.B.; Patterson, H.K.; Feng, J.; El Roubay, N.; Claudio, K.; Marcatto, L.R.; Tavares, L.C.; Galvez, J.M.; Calderon-Ospina, C.-A.; et al. Machine Learning for Prediction of Stable Warfarin Dose in US Latinos and Latin Americans. *Front. Pharmacol.* **2021**, *12*, 749786. [[CrossRef](#)] [[PubMed](#)]
45. Ioannidis, J.P.A.; Greenland, S.; Hlatky, M.A.; Khoury, M.J.; Macleod, M.R.; Moher, D.; Schulz, K.F.; Tibshirani, R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* **2014**, *383*, 166–175. [[CrossRef](#)]
46. van der Ploeg, T.; Austin, P.C.; Steyerberg, E.W. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.* **2014**, *14*, 137. [[CrossRef](#)] [[PubMed](#)]
47. Riley, R.D.; Ensor, J.; Snell, K.I.E.; Harrell, F.E., Jr.; Martin, G.P.; Reitsma, J.B.; Moons, K.G.M.; Collins, G.; van Smeden, M. Calculating the sample size required for developing a clinical prediction model. *BMJ* **2020**, *368*, m441. [[CrossRef](#)]
48. Sun, G.W.; Shook, T.L.; Kay, G.L. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J. Clin. Epidemiol.* **1996**, *49*, 907–916. [[CrossRef](#)]
49. Rosthøj, S.; Keiding, N. Explained variation and predictive accuracy in general parametric statistical models: The role of model misspecification. *Lifetime Data Anal.* **2004**, *10*, 461–472. [[CrossRef](#)]
50. Sheiner, L.B.; Beal, S.L. Some suggestions for measuring predictive performance. *J. Pharmacokinetic. Biopharm.* **1981**, *9*, 503–512. [[CrossRef](#)]
51. Gage, B.F.; Eby, C.; Johnson, J.A.; Deych, E.; Rieder, M.J.; Ridker, P.M.; Milligan, P.E.; Grice, G.; Lenzini, P.; Rettie, A.E.; et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin. Pharmacol. Ther.* **2008**, *84*, 326–331. [[CrossRef](#)] [[PubMed](#)]
52. James, A.H.; Britt, R.P. Prospective comparative study of computer programs used for management of warfarin. *J. Clin. Pathol.* **1993**, *46*, 781. [[CrossRef](#)] [[PubMed](#)]
53. Ageno, W.; Johnson, J.; Nowacki, B.; Turpie, A. A computer generated induction system for hospitalized patients starting on oral anticoagulant therapy. *Thromb. Haemost.* **2000**, *83*, 849–852. [[PubMed](#)]
54. Obermeyer, Z.; Emanuel, E.J. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *N. Engl. J. Med.* **2016**, *375*, 1216–1219. [[CrossRef](#)] [[PubMed](#)]
55. Verhoef, T.I.; Redekop, W.K.; Daly, A.K.; van Schie, R.M.F.; de Boer, A.; Maitland-van der Zee, A.-H. Pharmacogenetic-guided dosing of coumarin anticoagulants: Algorithms for warfarin, acenocoumarol and phenprocoumon. *Br. J. Clin. Pharmacol.* **2014**, *77*, 626–641. [[CrossRef](#)] [[PubMed](#)]
56. Flaker, G.; Lopes, R.D.; Hylek, E.; Wojdyla, D.M.; Thomas, L.; Al-Khatib, S.M.; Sullivan, R.M.; Hohnloser, S.H.; Garcia, D.; Hanna, M.; et al. Amiodarone, anticoagulation, and clinical events in patients with atrial fibrillation: Insights from the ARISTOTLE trial. *J. Am. Coll. Cardiol.* **2014**, *64*, 1541–1550. [[CrossRef](#)] [[PubMed](#)]
57. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Br. J. Surg.* **2015**, *102*, 148–158. [[CrossRef](#)]