ORIGINAL RESEARCH **Open access**

Explainable Neural-Symbolic Model for Clinical Decision Support Combining Deep Learning Predictions with Rule-Based Clinical Guidelines for Anticoagulation Management

Mateo Silva[1*], Bruno Santos[1], Ricardo Pinto[2], Luis Teixeira[1]

Abstract

Anticoagulation management requires balancing multiple factors such as bleeding risk, thromboembolic risk, drug interactions, and renal function. Deep learning can assist in risk prediction, but its effectiveness relies on clinicians' ability to understand and verify the recommendations. Black-box models may recommend actions without providing clear explanations. In contrast, clinical guidelines are rule-based but not directly executable by neural models. This article introduces a neuro-symbolic XAI framework that combines deep learning predictions with explicit clinical guidelines. It includes a neural prediction module, a symbolic reasoning engine, and an integration layer for traceable justifications. The neuro-symbolic approach connects data-driven predictions to clinical rules, improving auditability and trustworthiness in decision support. This framework aims to enhance anticoagulation management by providing verifiable, clinician-understandable decision support, focusing on explainability-by-design.

Keywords Clinical decision support, Warfarin dosing, Explainable artificial intelligence, Neuro-symbolic AI, Anticoagulation, Clinical guidelines

*Correspondence:
Mateo Silva
mateo.silva@gmail.com
[1] Department of Intelligent Healthcare Systems, University of Lisbon, Lisbon, Portugal
[2] Department of AI Clinical Engineering, University of Porto, Porto, Portugal

Introduction

Anticoagulation management is a high-stakes clinical problem because treatment decisions must continuously balance prevention of thromboembolism against avoidance of bleeding. Contemporary guideline documents for atrial fibrillation and venous thromboembolism provide structured recommendations for anticoagulant selection, monitoring, and patient risk stratification, yet these recommendations still require individualized interpretation in routine care [1-4]. Warfarin adds further complexity because its narrow therapeutic index, drug interactions, dietary sensitivity, and pharmacogenomic variability make dose selection highly patient-specific [5-8]. A clinical decision support system for this domain should therefore reason not only from predicted risks, but also from the explicit medical logic embedded in professional guidelines.

Deep learning and machine learning models could assist anticoagulation management by predicting bleeding risk, thromboembolic risk, discharge dosing needs, INR stability, or likely response to therapy. However, a clinically useful prediction is not the same as a clinically justified recommendation, especially when a clinician must understand why a model would favor dose reduction, drug continuation, closer monitoring, or reconsideration of therapy [9-12]. Studies of warfarin dosing and anticoagulation control illustrate how data-driven systems

can model complex patient-level relationships, but they also highlight the need for transparent reasoning when recommendations affect medication safety [13-17]. In high-stakes clinical AI, interpretability is not merely a usability feature; it is part of the safety argument for human oversight.

Neuro-symbolic AI offers a conceptual route for combining neural prediction with symbolic reasoning. Logic Tensor Networks, DeepProbLog, and broader neuro-symbolic approaches show how learned representations could interact with rules, probabilistic logic, or differentiable constraints rather than operating as isolated black boxes [18-21]. In treatment recommendation, Delphi-style neuro-symbolic causal reinforcement learning illustrates how symbolic constraints could guide proposed actions while retaining a structured basis for explanation [22]. These approaches are especially relevant to anticoagulation because guideline logic is already expressed in conditional clinical language that can, in principle, be encoded and audited.

The thesis of this article is that an explainable neuro-symbolic clinical decision support framework could combine deep learning predictions with executable anticoagulation guidelines to produce recommendations that are both patient-specific and clinically traceable. In this framework, the neural component would predict risk states or response tendencies, while the symbolic component would apply guideline-derived rules for drug choice, dose adjustment, monitoring intensity, and safety checks [2, 4, 23, 24]. The resulting system would not replace clinical judgment, but would present a recommendation together with a rule-level explanation trace. This roadmap emphasizes conceptual architecture, implementation logic, and validation requirements rather than empirical performance claims.

Background

Anticoagulation clinical practice guidelines

Clinical practice guidelines for anticoagulation formalize medical knowledge into structured recommendations concerning atrial fibrillation, venous thromboembolism, stroke prevention, bleeding risk management, and therapy monitoring. The ESC, ACC/AHA/ACCP/HRS, and CHEST guidelines provide complementary decision logic for anticoagulant use, including when therapy would be considered, how clinical risk factors should be interpreted, and when monitoring or reassessment would be appropriate [1-4, 25]. For warfarin, clinical reasoning must also account for pharmacogenomic factors and patient-specific covariates that influence dose response, which makes guideline-compatible personalization especially important [6-8]. In a neuro-symbolic system, these sources of authoritative knowledge would be translated into executable rules while preserving their clinical meaning.

Deep learning in anticoagulation

Deep learning and machine learning could support anticoagulation care by producing patient-specific predictions related to bleeding risk, thromboembolic risk, INR stability, warfarin dose response, or medication management trajectories. Existing work on warfarin dosing, anticoagulation control, and discharge dose prediction illustrates the range of predictive tasks that could supply inputs to a decision support framework [5, 13-16, 26]. Related models for longitudinal warfarin levels and individualized treatment strategies show how temporal patient data could inform dynamic anticoagulation decisions without directly replacing clinical rules [27, 28]. In the proposed framework, such models would serve as risk-estimation components whose outputs are passed to symbolic guideline execution rather than treated as final recommendations.

Explainable AI in anticoagulation

Explainable AI methods such as SHAP, LIME, feature attribution, attention visualization, and interpretable model design can help clinicians inspect which patient variables may influence a prediction. In pharmacogenomic and anticoagulation contexts, XAI could be used to explain why variables such as genotype, comedications, age, renal function, prior bleeding, or clinical instability would alter a predicted response or risk profile [6, 8, 29, 30]. However, post-hoc explanation alone does not ensure that a model reasons according to accepted clinical guidelines, because an attribution map may describe the prediction without proving that the recommendation follows clinical rules [10-12]. A neuro-symbolic approach therefore shifts the explanatory target from “why did the model predict this?” to “how did the prediction interact with the guideline logic to produce this recommendation?”

Neuro-symbolic AI in healthcare

Neuro-symbolic AI combines neural representation learning with symbolic structures such as rules, logic programs, constraints, and causal decision processes. Logic Tensor Networks and DeepProbLog exemplify how symbolic knowledge could be integrated with differentiable learning or probabilistic reasoning, while broader neuro-symbolic research frames this integration as a way to improve reasoning, interpretability, and knowledge use [18-21, 31]. In healthcare, such architectures are conceptually attractive because clinical knowledge is often codified as guidelines, eligibility criteria, contraindications, and conditional safety rules [23, 24]. Delphi further illustrates how a treatment recommendation framework could combine learning-based estimation with symbolic clinical constraints to support interpretable decision pathways [22].

Integration of guidelines as executable rules

Encoding clinical practice guidelines as executable rules would involve translating narrative recommendations into conditional logic, probabilistic clauses, priority rules, or fuzzy thresholds that can operate on structured patient data. For anticoagulation, such rules could represent drug eligibility, contraindications, monitoring triggers, dose adjustment conditions, pharmacogenomic considerations, and escalation pathways derived from clinical guidance [1-4]. Symbolic execution could then provide a transparent record of which conditions were satisfied, which exceptions were activated, and which recommendation pathway was selected [18, 19, 24]. This differs from ordinary post-hoc XAI because the explanation is generated by the actual decision process rather than appended after the prediction.

Framework Overview

High-level architecture

The proposed architecture begins with structured patient data from the electronic health record, including demographic variables, diagnoses, medications, laboratory values, renal function, prior bleeding history, thromboembolic risk factors, and available pharmacogenomic information. A neural prediction module would transform these data into structured risk estimates, such as bleeding risk tendency, thromboembolic risk tendency, INR instability likelihood, or warfarin response profile, drawing conceptually on predictive modeling work in anticoagulation and clinical AI [13-17]. A symbolic rule module would then execute guideline-derived logic over these estimates and patient facts, while the explainable integration layer would generate a decision trace connecting predictions, rules, and recommendations [2, 18, 19, 22]. The final output would be a clinician-facing recommendation with an auditable explanation rather than an opaque model score.

Figure 1 presents the proposed explainable neuro-symbolic architecture linking patient-specific anticoagulation evidence, neural risk estimation, executable guideline reasoning, auditable explanation traces, and clinician-supervised decision support.

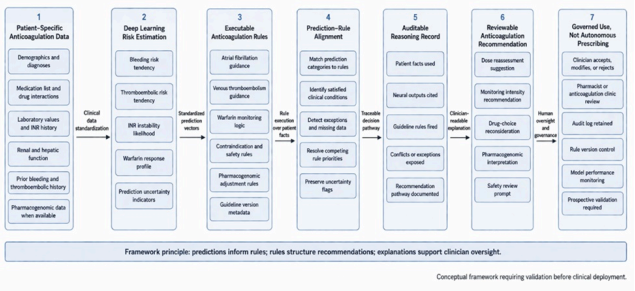

Figure 1. Explainable Neuro-Symbolic Architecture for Guideline-Concordant Anticoagulation Decision Support

Core assumptions

The framework assumes that relevant clinical practice guidelines can be digitized into a machine-readable representation without distorting their clinical intent. It also assumes that the clinical site has access to sufficiently structured patient data to support neural prediction and symbolic rule execution, including medication records, laboratory values, diagnostic codes, and relevant clinical history [23, 24]. Because anticoagulation decisions are safety-critical, the framework further assumes that local regulatory and institutional governance would permit AI-assisted recommendations only under clinician oversight [11, 24, 32]. These assumptions are not trivial, and each would require validation, governance review, and workflow testing before implementation.

3.3 Design principles

The first design principle is explainability-by-design, meaning that recommendations should be generated through transparent interaction between neural predictions and symbolic rules rather than explained only after the fact. The second principle is auditability, because a clinician or

reviewer should be able to inspect which guideline rules fired, which patient facts supported them, and how conflicts were resolved [10, 12, 24]. The third principle is modularity, where predictive models and guideline rules can be updated through separate pathways so that a guideline revision does not necessarily require retraining a neural model [18-20]. The fourth principle is clinical conservatism, because high-stakes medication decisions should prefer justifiable, reviewable recommendations over opaque automation [32].

Table 1 clarifies how the proposed framework separates prediction, rule execution, explanation generation, clinical presentation, and governance so that anticoagulation recommendations remain inspectable rather than opaque.

Table 1. Functional Separation of Neural, Symbolic, and Integration Components in Explainable Anticoagulation Decision Support

Framework component	Primary function	Clinical knowledge represented	Mair
Structured patient evidence layer	Converts EHR, laboratory, medication, and history data into computable inputs	Patient-specific anticoagulation context	Stan clinic
Neural prediction module	Estimates patient-specific risk and response tendencies	Learned statistical patterns from clinical data	Ble ten thromb tende ins wa respor
Symbolic guideline module	Executes encoded anticoagulation rules	Clinical practice guidelines, contraindications, monitoring logic, pharmacogenomic rules	Fire block exc priority
Neuro-symbolic integration layer	Aligns prediction categories with executable clinical rules	Relationship between predicted risk states and guideline conditions	Ma predic pa
Explanation trace generator	Records the reasoning pathway	Patient facts, prediction values, rule activations, conflicts, uncertainty	Au decis
Clinician-facing output layer	Presents reviewable decision support	Action framing under human supervision	Dose mo recomm drug recons safet
Governance and maintenance layer	Tracks model, rule, and guideline versions	Institutional policy, audit logs, update history	Ver revi syst

Neural Prediction Module

Patient-specific risk predictions

The neural prediction module would estimate patient-specific risk states or response profiles that are clinically relevant to anticoagulation management. These predictions could include bleeding tendency, thromboembolic tendency, INR instability, warfarin sensitivity, discharge dose needs, or likely pharmacogenomic response, depending on available patient data and local clinical use cases [5-8, 13, 14]. Reinforcement learning concepts could also inform dynamic treatment planning, where the system would reason about sequential dose adjustment or monitoring decisions, but such use would require careful clinical validation and safety constraints [22, 27, 33]. The module's purpose is not to issue the final recommendation, but to provide structured predictive evidence for the symbolic reasoning engine.

Output standardisation

For integration with symbolic reasoning, neural outputs must be standardized into clinically meaningful prediction vectors rather than unstructured scores. For example, the module could output normalized representations of bleeding risk, stroke risk, INR instability, drug response tendency, and uncertainty indicators that symbolic rules can interpret consistently [15-17, 26, 28]. This standardization

would make it possible for a rule engine to evaluate conditions such as elevated bleeding concern, unstable anticoagulation control, or pharmacogenomic sensitivity without depending on model-specific internals. It would also support explanation because each rule trigger could cite the relevant prediction category and patient evidence rather than merely referencing a black-box model output [9, 10, 12].

Symbolic Rule Module

Encoding clinical guidelines as executable rules

The symbolic rule module would convert anticoagulation guideline statements into executable clinical logic using first-order rules, probabilistic logic, differentiable logic, or related symbolic representations. A rule might state that if predicted bleeding concern is elevated, prior bleeding is present, renal function is impaired, and the guideline-defined clinical context supports reassessment, then the system should recommend closer review, modified monitoring, or reconsideration of therapy under clinician supervision [1-4]. Logic Tensor Networks and DeepProbLog provide conceptual templates for combining rule structures with learned information, while guideline-aware clinical decision support emphasizes the need for responsible, inspectable recommendations [7, 8, 26, 27]. Importantly, such rules would be illustrative and would require formal clinical review before use.

Integration of pharmacogenomic knowledge

Pharmacogenomic reasoning is especially relevant for warfarin because genotype-informed sensitivity can influence how a dose recommendation is interpreted. The symbolic module could encode gene-drug interaction rules involving factors such as VKORC1, CYP2C9, and CYP4F2, while the neural module could provide a patient-specific response tendency based on structured clinical and genetic information [6-8, 29, 30]. In this design, pharmacogenomic knowledge would not be hidden inside a model; it would be represented as explicit clinical logic that can be inspected, updated, and linked to the recommendation trace. Such a module could support personalized dosing discussions while preserving the distinction between predictive evidence and guideline-authorized action.

Handling conflicting and conditional rules

Anticoagulation guidelines often contain conditional recommendations, exceptions, and competing safety considerations, so the symbolic rule module must include mechanisms for conflict handling. Priority ordering, fuzzy logic, weighted rules, defeasible reasoning, or probabilistic symbolic methods could be used to represent situations in which bleeding concern, thromboembolic concern, renal impairment, drug interactions, and patient context point toward different actions [2-4, 6]. For example, a rule supporting anticoagulation continuation could conflict with a safety rule requiring reassessment, and the system would need to expose that conflict rather than silently resolving it. Guideline versioning would also be essential, because updated ACC/AHA, ESC, or CHEST recommendations should be traceable to the exact rule set used at the time of decision support [1-4, 25].

Explainable Integration Layer

Decision trace generation

The explainable integration layer would record how neural predictions and symbolic rules interact to produce an anticoagulation recommendation. For each recommendation, the system could document the relevant patient facts, the prediction categories supplied by the neural module, the guideline-derived rules that fired, and the conflict-resolution logic applied by the symbolic engine [18, 19, 22]. This trace would function as the internal explanation of the decision process, not as a post-hoc approximation of a black-box output. Such a design aligns with the broader XAI argument that clinical explanations should support clinician review, safety reasoning, and accountable oversight in high-stakes settings [10, 11, 24].

Explanation presentation

Explanation presentation would translate the decision trace into clinician-readable language while preserving access to the underlying rule logic. A recommendation could be accompanied by a concise statement that identifies the predicted risk category, the guideline condition that was satisfied, the patient factors that activated the rule, and any unresolved uncertainty requiring clinician judgment [12, 23, 24]. Attention-based explanation methods may be useful for highlighting influential clinical inputs, but attention alone should not be treated as a complete justification unless it is

connected to explicit rule execution and clinical reasoning [34]. The goal is not to make the system persuasive, but to make its reasoning inspectable, challengeable, and clinically reviewable.

Clinical Workflow Integration

Deployment as decision support

In clinical workflow, the proposed system would operate as a decision support assistant rather than an autonomous prescriber. It could propose actions such as dose reassessment, monitoring review, drug-choice reconsideration, or pharmacogenomic interpretation, while the clinician would remain responsible for judging patient context, preferences, contraindications, and institutional policy [23, 24, 32]. The explanation trace would allow the clinician to inspect whether the recommendation followed the expected guideline pathway and whether the neural prediction appeared clinically plausible [9, 10]. This design is especially important for anticoagulation because medication safety depends on contextual judgment as well as structured risk assessment.

Update and maintenance

The framework would support separate update pathways for predictive models and symbolic guideline rules. When anticoagulation guidelines are revised, the symbolic rule base could be re-encoded, versioned, and reviewed without necessarily retraining the neural prediction module, while population shifts or local practice changes could trigger separate model reassessment [1-4]. This separation would make maintenance more transparent because a recommendation trace could identify which guideline version and which prediction module version were active at the time of decision support [24]. It would also reduce the risk that clinical knowledge updates become hidden inside opaque model retraining procedures.

Evaluation Strategy

Explainability metrics

Evaluation should examine whether clinicians can understand, verify, and challenge the recommendation pathway produced by the neuro-symbolic system. Conceptual explainability criteria could include trace completeness, rule-execution correctness, clinical plausibility, consistency between patient facts and fired rules, and the usefulness of explanations for shared decision review [10-12]. The evaluation should also compare explanation types, distinguishing feature attribution, attention visualization, symbolic rule traces, and integrated neuro-symbolic justifications [9, 34]. Because this article is framework-oriented, such criteria are presented as validation requirements rather than as reported performance findings.

Table 2 provides a validation matrix showing that the framework must be assessed not only for predictive performance, but also for guideline fidelity, explanation quality, safety behavior, workflow fit, and long-term governance.

Table 2. Validation Matrix for Explainable Neuro-Symbolic Anticoagulation Decision Support

Validation domain	Core question	What should be evaluated	Evide requi befo deploy
Predictive validity	Are neural risk estimates clinically reliable?	Bleeding risk, thromboembolic risk, INR instability, dose-response prediction, uncertainty calibration	Retrosp an prospe validati loc anticoag popula
Guideline fidelity	Are symbolic rules faithful to clinical guidance?	Rule encoding, exceptions, contraindications, monitoring triggers, pharmacogenomic logic	Expert r by clini pharma informat and gui specia
Neuro-symbolic alignment	Do predictions activate rules appropriately?	Mapping between risk categories and executable rule conditions	Scen testing a typic borderlir confli cas
Explanation quality	Can clinicians understand and challenge	Trace completeness, readability, patient-factor	Hum cente usab testing

	the reasoning pathway?	visibility, rule transparency, uncertainty communication	anticoag clinici
Safety behavior	Does the system behave conservatively under uncertainty?	Missing data, conflicting rules, edge cases, renal impairment, drug interactions, high-risk bleeding contexts	Stress t simulatic safety- revi
Workflow fit	Does the tool support clinical work without replacing judgment?	Timing, alert burden, pharmacist review, clinician override, documentation burden	Pilot wo evaluat anticoag clinic hosp setti
Governance readiness	Can the system be maintained over time?	Rule versioning, model monitoring, audit logs, update approval, retirement of outdated rules	Institut govern protoco mainte pla

Safety and consistency

Safety evaluation would test whether the symbolic module behaves consistently at rule boundaries, under missing data, and in clinically ambiguous cases. Borderline scenarios could be simulated conceptually to assess whether the system exposes uncertainty, detects contradictory rules, and avoids unjustified escalation from prediction to recommendation [18, 19, 31]. Reinforcement learning-inspired dosing or treatment sequencing would require especially careful constraint checking because sequential recommendations may create downstream safety implications [22, 27, 33]. In this framework, safety depends not only on predictive quality but also on logical consistency, traceability, and clinician control.

Demonstrator use case

A conceptual demonstrator could focus on warfarin anticoagulation management because warfarin combines narrow therapeutic range, pharmacogenomic variability, longitudinal monitoring, and frequent dose reassessment. The neural module would predict tendencies such as INR instability or warfarin response, while the symbolic module would encode clinical and pharmacogenomic adjustment logic involving patient factors, comedications, and genotype-related sensitivity [5-8, 26]. The integration layer would then show why a dose review, closer monitoring action, or clinician reassessment was suggested, linking the recommendation to both predicted risk and explicit rule activation [28-30]. Such a demonstrator would remain illustrative until validated in a governed clinical environment.

Guideline-concordance review

A separate evaluation pathway should assess whether the system's recommendations remain concordant with accepted anticoagulation guidance across common clinical scenarios. Expert reviewers could inspect whether rules derived from atrial fibrillation and venous thromboembolism guidance are represented faithfully, whether exceptions are visible, and whether the system distinguishes strong guideline logic from contextual clinical judgment [1-4, 25]. This review would be particularly important when symbolic rules simplify narrative guideline text into executable conditions. The evaluation should therefore treat guideline encoding itself as a safety-critical artifact requiring clinical, technical, and governance review.

Human-centered explanation review

Human-centered review would examine whether clinicians can use the explanation trace without becoming overloaded by technical detail. The system should present a summary explanation first, followed by expandable details on prediction inputs, rule activation, guideline source, conflict handling, and uncertainty [10, 11, 24]. Clinical AI examples from other domains show that deep learning can be made more clinically actionable when outputs are embedded in interpretable referral or decision pathways, but anticoagulation would require its own validation because medication recommendations involve different risks and responsibilities [35]. The key evaluation question is whether explanation improves safe review, not whether it merely increases user confidence.

Limitations

Technical limitations

A major technical limitation is rule explosion, because anticoagulation guidelines contain many conditional

statements, exceptions, comorbidities, and context-dependent judgments. As the symbolic rule base grows, interactions among rules may become difficult to verify, especially when neural prediction categories are uncertain or when multiple guideline pathways apply simultaneously [18-20]. Formal verification could reduce contradiction risk, but it may not fully capture clinical nuance or local policy variation [31]. The framework therefore requires careful rule governance, modular testing, and continuous review rather than assuming that symbolic logic automatically guarantees safety.

Clinical limitations

Clinically, the proposed system should be understood as an assistant that supports review rather than an autonomous decision-maker. Anticoagulation decisions involve patient preferences, clinician judgment, institutional protocols, bleeding history, procedural plans, drug availability, and shared decision-making considerations that may not be completely captured by structured data or guideline rules [23, 24, 32]. XAI can improve inspectability, but it cannot eliminate liability, accountability, or regulatory questions surrounding AI-assisted medication management [9]. Any deployment would therefore require institutional oversight, clinician training, and prospective validation in the intended clinical setting.

Implementation challenges

Implementation would require structured digitized guidelines, reliable EHR integration, consistent terminology mapping, and maintenance workflows for rule version control. Many institutions lack machine-readable anticoagulation pathways that can be directly translated into executable rules, and vendor-dependent EHR architectures may complicate real-time integration [23, 24]. Pharmacogenomic data may also be unavailable, inconsistently documented, or difficult to connect to medication decision support workflows [6-8]. These challenges mean that the framework is not simply a modeling problem, but a sociotechnical implementation project.

Data and representation limits

The neural module would depend on structured patient data, yet anticoagulation decisions often require information contained in free text, external records, patient reports, or evolving clinical circumstances. Missing laboratory values, incomplete medication reconciliation, uncertain adherence, undocumented bleeding events, or unavailable genotype information could all affect the reliability of predicted risk states [13-17]. The symbolic module would also need explicit rules for missingness, uncertainty, and deferral to clinician review. Without these safeguards, the system could appear more certain than the underlying clinical data justify.

Explainability limits

Although the framework is designed for explainability, explanations can still be incomplete, misleading, or too complex for routine use. Feature attribution methods may identify influential variables without proving causal relevance, while symbolic traces may show rule activation without fully explaining why the underlying prediction took its value [9, 10, 12]. Attention-based displays are also limited when used as standalone explanations, because highlighted inputs do not necessarily constitute clinical reasoning [34]. A robust XAI design should therefore combine attribution, uncertainty communication, rule traces, and clinician review rather than relying on any single explanatory device.

Governance and maintenance limits

Long-term governance would be necessary because clinical guidelines, drug labeling, local protocols, and patient populations change over time. A neuro-symbolic anticoagulation system would require versioned rule repositories, documentation of model updates, audit logs, and processes for retiring outdated rules [1, 2, 4, 24]. This maintenance burden may be substantial, but it is also the source of the framework's transparency because changes to clinical knowledge can be inspected rather than hidden inside an opaque model revision. The practical value of the system would therefore depend on sustained collaboration among clinicians, informaticians, AI developers, pharmacists, and institutional governance bodies.

Conclusion

The proposed neuro-symbolic framework integrates neural prediction with symbolic guideline execution for explainable anticoagulation decision support. The neural component would estimate patient-specific risks and response tendencies, while the symbolic component would apply executable clinical rules derived from anticoagulation guidance.

Its main advantage is explainability-by-design. Rather than producing a recommendation first and explaining it afterward, the framework would generate a traceable pathway from patient data to prediction, from prediction to rule activation, and from rule activation to clinician-facing recommendation.

Important challenges remain. Rule complexity, formal verification, clinical nuance, regulatory acceptance, data quality, and long-term maintenance would all need careful attention before such a framework could be used responsibly in clinical environments.

Future development should involve close collaboration among guideline committees, anticoagulation clinicians, pharmacists, informaticians, AI researchers, regulators, and implementation teams. Carefully governed pilots in anticoagulation clinics could help determine how neuro-symbolic decision support might become trustworthy, auditable, and clinically useful.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Jan 2026 Revised: 19 Jan 2026 Accepted: 18 Feb 2026
Published online: 20 July 2026

References

Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J. 2021;42(5):373-498.

Members WC, Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2024;149(1):e1-e156.

Lip GY, Banerjee A, Boriani G, Chiang CE, Fargo R, Freedman B, et al. Antithrombotic therapy for atrial fibrillation: CHEST guideline and expert panel report. Chest. 2018;154(5):1121-201.

Stevens SM, Woller SC, Kreuziger LB, Bounameaux H, Doerschug K, Geersing GJ, et al. Antithrombotic therapy for VTE disease: second update of the CHEST guideline and expert panel report. Chest. 2021;160(6):e545-608.

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(10):e0205872. https://doi.org/10.1371/journal.pone.0205872.

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. 2020;10:1550. https://doi.org/10.3389/fphar.2019.01550.

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. https://doi.org/10.1002/psp4.12721.

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. https://doi.org/10.1002/cpt.2360.

Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2(10):749-60. https://doi.org/10.1038/s41551-018-0304-0.

Yang CC. Explainable artificial intelligence for predictive modeling in healthcare. J Healthc Inform Res. 2022;6(2):228-39. https://doi.org/10.1007/s41666-022-00114-0.

Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. https://doi.org/10.1186/s12911-020-01332-6.

Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst. 2021;32(11):4793-813. https://doi.org/10.1109/TNNLS.2020.3027314.

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. https://doi.org/10.1038/s41598-024-54873-4.

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. https://doi.org/10.1016/j.compbiomed.2022.106548.

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. https://doi.org/10.1038/s41598-023-49667-2.

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. https://doi.org/10.2196/47262.

Gordon J, Norman M, Hurst M, Mason T, Dickerson C, Sandler B, et al. Using machine learning to predict anticoagulation control in atrial fibrillation: a UK Clinical Practice Research Datalink study. Inform Med Unlocked. 2021;25:100688. https://doi.org/10.1016/j.imu.2021.100688.

Badreddine S, Garcez AD, Serafini L, Spranger M. Logic tensor networks. Artif Intell. 2022;303:103649. https://doi.org/10.1016/j.artint.2021.103649.

Manhaeve R, Dumančić S, Kimmig A, Demeester T, De Raedt L. Neural probabilistic logic programming in DeepProbLog. Artif Intell. 2021;298:103504. https://doi.org/10.1016/j.artint.2021.103504.

Garcez AD, Lamb LC. Neurosymbolic AI: the 3rd wave. Artif Intell Rev. 2023;56(11):12387-406. https://doi.org/10.1007/s10462-023-10448-w.

Donadello I, Serafini L, Garcez AD. Logic tensor networks for semantic image interpretation. arXiv [Preprint]. 2017:arXiv:1705.08968.

Tao M, Qin H, Fang Y, Shan C, Tan T. Delphi: a neuro-symbolic framework for individualized, safe and interpretable treatment recommendation. Proc AAAI Conf Artif Intell. 2026;40(23):19389-97.

Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199-200. https://doi.org/10.1001/jama.2018.17163.

Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inform Assoc. 2024;31(11):2730-9.

January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. J Am Coll Cardiol. 2019;74(1):104-32. https://doi.org/10.1016/j.jacc.2019.01.011.

Ahn S. Building and analyzing machine learning-based warfarin dose prediction models using scikit-learn. Transl Clin Pharmacol. 2022;30(4):172-9. https://doi.org/10.12793/tcp.2022.30.e19.

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 Inform Assoc. 2022;29(10):1722-32.

Xue L, He S, Singla RK, Qin Q, 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. 2024;110(10):6528-40. https://doi.org/10.1097/JS9.0000000000001814.

Martínez-Jiménez JE, Ortega-Lampón Y, Cedres-Rivera D, Heredia-Negrón F, Roche-Lima A, Duconge J. Leveraging machine learning to predict warfarin sensitivity in the Puerto Rican population: a pharmacogenomic approach. Int J Environ Res Public Health. 2026;23(3):337.

Zhong C, Ye ZH, Xiang GU. Application of an interpretable neural network framework based on the LASSO-proj algorithm for warfarin dose prediction. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi. 2025;42(6):1235.

Yang XW, Shao JJ, Guo LZ, Zhang BW, Zhou Z, Jia LH, et al. Neuro-symbolic artificial intelligence: towards improving the reasoning abilities of large language models. arXiv [Preprint]. 2025:arXiv:2508.13678.

Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-15. https://doi.org/10.1038/s42256-019-0048-x.

Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(11):1716-20. https://doi.org/10.1038/s41591-018-0213-5.

Jain S, Wallace BC. Attention is not explanation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019. p. 3543-56.

De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342-50. https://doi.org/10.1038/s41591-018-0107-6.