New AI Prediction Model Using Serial PT-INR Measurements in AF Patients on VKAs: GARFIELD-AF

Shinichi Goto[1], Shinya Goto[2], Karen S. Pieper[3], Jean-Pierre Bassand[4], A. John Camm[5], David A. Fitzmaurice[6], Samuel Z. Goldhaber[7], Sylvia Haas[8], Alexander Parkhomenko[9], Ali Oto[10], Frank Misselwitz[11], Alexander G.G. Turpie[12], Freek W.A. Verheugt[13], Keith A.A. Fox[14], Bernard J. Gersh[15], Ajay K. Kakkar[16]; for the GARFIELD-AF Investigators

[1]Department of Cardiology, Keio University School of Medicine, Tokyo, Japan; [2]Department of Medicine (Cardiology), Tokai University School of Medicine, Isehara, Japan; [3]Thrombosis Research Institute, London, UK; [4]Thrombosis Research Institute, London, UK and University of Besançon, Besançon, France; [5]Cardiology Clinical Academic Group, Molecular & Clinical Sciences Institute, St. George's University of London, London, UK; [6]Department of Cardio-respiratory Primary Care, Warwick Medical School, University of Warwick, Coventry, UK; [7]Harvard Medical School, Boston, MA, USA; [8]Formerly Klinikum rechts der Isar, Technical University of Munich, Munich, Germany; [9]National Scientific Center, MD Strazhesko Institute of Cardiology, Kiev, Ukraine; [10]Chairman, Department of Cardiology, MHG, Memorial Ankara Hospital, Ankara, Turkey; [11]Bayer AG, Pharmaceuticals Division, Berlin, Germany; [12]McMaster University, Hamilton, Ontario, Canada; [13]Department of Cardiology, Onze Lieve Vrouwe Gasthuis (OLVG), Amsterdam, Netherlands; [14]Edinburgh Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK; [15]Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; [16]Thrombosis Research Institute and University College London, London, UK

Brief Title: New AI Prediction Model in AF

Address for correspondence

Shinya Goto, MD, PhD

Department of Medicine (Cardiology), Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

Tel.: +81-463-93-1121; e-mail: sgoto3@mac.com

Abstract

Background Most clinical risk stratification models are based on measurement at a single time-point rather than serial measurements. Artificial intelligence (AI) is able to predict one-dimensional outcomes from multi-dimensional datasets. Using data from GARFIELD-AF registry, a new AI model was developed for predicting clinical outcomes in atrial fibrillation (AF) patients up to 1 year based on sequential measures of PT-INR within 30 days of enrolment.

Methods and results Patients with newly diagnosed AF who were treated with vitamin K antagonists (VKA) and had at least 3 measurements of PT-INR taken over the first 30 days after prescription were analyzed. The AI model was constructed with multilayer neural network including long short-term memory (LSTM) and one-dimensional convolution layers. The neural network was trained using PT-INR measurements within days 0–30 after starting treatment and clinical outcomes over days 31–365 in a derivation cohort (cohorts 1–3; n = 3185). Accuracy of the AI model at predicting major bleed, stroke/SE, and death was assessed in a validation cohort (cohorts 4–5; n = 1523). The model's c-statistic for predicting major bleed, stroke/SE, and all-cause death was 0.75, 0.70, and 0.61, respectively.

Conclusions Using serial PT-INR values collected within 1 month after starting VKA, the new AI model performed better than time in therapeutic range (TTR) at predicting clinical outcomes occurring up to 12 months thereafter. Serial PT-INR values contain important information that can be analyzed by computer to help predict adverse clinical outcomes.

Keywords atrial fibrillation (AF), artificial intelligence (AI), machine learning

Introduction

In chronic diseases such as atrial fibrillation (AF) risk stratification using prediction models is useful for clinical decision making. Several models predict clinical events such as stroke and bleeding.[1–3] The CHA_2DS_2-VASc and HAS-BLED scores are widely used to select suitable AF patients for oral anticoagulation (OAC).[4–6] However, some of the variables in these scoring systems are not consistently related to outcomes.[7] Novel machine learning technology has facilitated the development of more accurate models such as the GARFIELD-AF risk model.[8] However, these models incorporate data obtained at a single time-point, baseline. Although computers can process multi-dimensional data such as changes of variables over time, few models have used these inputs to predict future clinical events.[9,10]

Vitamin K antagonists (VKAs) continue to be prescribed for the prevention of stroke in patients with AF, despite the more recent introduction of non-VKA oral anticoagulants (NOACs).[11,12] VKAs are the only recommended choice of OAC for AF patients with hemodynamically overt mitral stenosis and mechanical heart valve. Clinicians adjust the dose of VKA based on an individual patient's prothrombin time international normalized ratio (PT-INR) at each visit. Time in therapeutic range (TTR) is widely used to standardize the effects of VKA therapy over periods beyond 6 months.[13–17] Various bleeding risk scores feature a TTR component to enhance accuracy,[18] and TTR has predictive power for thrombotic and bleeding events.[19,20] However, information on serial changes in PT-INR during early-phase VKA therapy, which may reflect many occult clinical characteristics of patients such as genotype,[21,22] concomitant medications,[23] and lifestyle,[24] were not included in these TTR-based models.

Advances in artificial intelligence (AI) using recurrent neural networks (RNN) allow the identification and translation of multi-dimensional data including time-series data directly into meaningful models.[25] Herein, we describe a new AI model for predicting clinical outcomes over 31–365 days after patient enrolment. The model evaluates serially measured PT-INR within the first 30 days of treatment only without other clinical parameters, using data from the largest multinational prospective registry in AF, Global Anticoagulant Registry in the Field (GARFIELD)-AF. The predictive accuracy of the AI model was compared with that of TTR. The working hypothesis was to test whether serially measured PT-INR in early phase can provide information to predict future clinical events.

Methods

Design

The AI model was derived from prospective GARFIELD-AF data gathered in adults with newly diagnosed AF.[26] Three independent AI models were developed with the same composite of neural network structure with multi-dimensional patient-level PT-INR values obtained within the first 30 days after starting treatment. The model tabulated the clinical events of major bleed, ischemic stroke/systemic embolism (SE), and death occurring within days 31–365.

Registry population

GARFIELD-AF is an ongoing, international, prospective registry of newly diagnosed patients with AF at risk of stroke. The study design, baseline characteristics, and main results have been published.[26–29] Eligible patients were adults aged >18 years who had

been diagnosed with nonvalvular AF within the previous 6 weeks and had at least one risk factor for stroke as judged by the investigator. Risk factors were not prespecified in the protocol. Any use of antithrombotic agents was shared decision between clinicians and patients only. Patients with a transient reversible cause of AF and those for whom follow-up was not envisaged were excluded. The present analysis was conducted in patients enrolled in GARFIELD-AF cohorts 1–5 between March 2010 and August 2016. Data were extracted from the study database in November 2017.

Study population

Patients who received anticoagulation therapy with VKA and had 3 or more PT-INR measurements within the first 30 days after enrolment were included in the model. Patients were excluded if they had experienced any outcome events such as serious bleeding or stroke or died within the first 30 days. In this analysis day of first visit was set as day 0. Patients from cohorts 1–3 (recruited between March 2010 and October 2014) were included in the derivation cohort whereas those in cohort 4–5 (recruited March 2014–August 2016) comprised the validation cohort. This study design was considered stringent because each GARFIELD-AF cohort exhibited substantial differences in terms of participating countries, use of anticoagulants, and outcomes.[30]

Follow-up

Collection of follow-up data occurred at 4-monthly intervals based on medical records and, sometimes, telephone interviews up to 24 months. The incidence of ischemic stroke, transient ischemic attack (TIA), SE, acute coronary syndrome (ACS), hospitalization, death (cardiovascular and non-cardiovascular), chronic heart failure

(CHF; occurrence or worsening), and bleeding (severity and location) was documented. An audit and quality control programme was applied, and data were examined for completeness and accuracy by the coordinating centre (TRI, London, UK). By design, 20% of all electronic case report forms in the GARFIELD-AF registry were monitored against source documentation at sites over the 8 years of recruitment and follow-up.

Outcomes

Outcome measures used in this analysis were major bleeding, stroke/SE, and all-cause death occurring between days 31 and 365. Major bleed was classified by investigators according to International Society on Thrombosis and Haemostasis definition.[31] Stroke/SE was defined as a combined end point of ischemic stroke, SE, and TIA.

Artificial intelligence model

The structure of neural networks for the AI model is shown in Figure 1A. To deal with serial data on raw PT-INR measurements, the AI model was constructed by stacking multiple layers of special neurons that can deal with time-dependent data, namely one-dimensional convolution layer and long short-term memory (LSTM) layer. The LSTM layer transfer rectified data to each neighboring neuron.[32] This structure allows the layer to learn time-dependent data in sequential order.

The neural network model was trained independently for each outcome event. For training, PT-INR measurement patterns for each individual patient were converted to a 30-dimension PT-INR vector as shown in Figure 1B. All PT-INR measurements obtained within the first 30 days were input to the model. The measured PT-INR

value was inserted into nth element of the 30-dimension vector, where n is the number of days after starting VKA. Unmeasured data-points were filled with 0. Each vector for patients was labelled with the occurrence of outcome (0 for no event and 1 for event for all three outcome measures) within days 31–365. The neural networks were trained with the multi-dimensional dataset of the PT-INR vector and outcome label as shown in Figure 1C.

Model training

The process of model training is shown in Figure 2. Training was performed using only patient data from the derivation cohort. The derivation cohort was further split into training (70%) and testing (30%) datasets. Training was performed for 500 epochs and each training epoch included a mini-batch of 455 patients randomly selected from the training dataset. Conceptually, performance of the model is designed to improve by training with longer epochs. However, this approach can also result in overfitting. To avoid this pitfall and select the model with best performance, the model was evaluated using the testing dataset at the end of each epoch. The final model was that which performed best with the testing dataset. Performance was measured by calculating the c-statistics of the prediction model for all the data in testing dataset. No data from validation cohort were used for training.

Model validation

The derived models were validated by inputting the 30-day PT-INR vector and obtaining prediction scores for each outcome. Predicted outcomes were compared with the actual clinical course for each individual patient in the validation cohort. Receiver operating characteristic (ROC) curves were drawn to evaluate the predictive

value of the model. The threshold to achieve overall best accuracy for the model was determined and the model's sensitivity and specificity calculated at that threshold. To test the ability of the model to discriminate between high- and low-risk patients for each event, three sets of Kaplan-Meier plots were drawn for event rates stratified as high and low risk with the threshold.

Ethics

All trial protocols were approved by independent ethics committee and hospital-based institutional review board. The registry was conducted in accordance with the Declaration of Helsinki, local regulatory requirements, and International Conference on Harmonisation-Good Pharmacoepidemiological and Clinical Practice guidelines. All patients provided written informed consent to participate.

Statistical analysis

The neural network was constructed and trained using Keras framework version 2.1.6 (https://keras.io) and TensorFlow version 1.8.0 as backend.[33] The neural network was trained using the back-propagation supervised training algorithm. The loss function that was previously reported to reflect the c-statistics[34] was minimized using the RMSprop optimizer.

The c-statistics, the threshold to achieve best accuracy and corresponding accuracy, sensitivity, and specificity of the model at the threshold with its 95% confidence intervals (95% CI) were calculated by bootstrap procedure with 2000 bootstrap rounds using the pROC package of R version 3.5.1.[35] Comparison of ROC curve between the AI model and TTR was performed similarly with pROC package. Kaplan-Meier plots

were produced using the survival package of R. All p-values were calculated by log-rank test; a p-value <0.05 was considered statistically significant.

Results

Patients

The flowchart of patient selection is shown in Figure 3. Of 14,437 de novo AF patients treated with VKA, 4806 had at least 3 PT-INR measurements within the first 30 days and were included in the analysis. Ninety-eight patients were excluded (92 with an outcome event within the first 30 days and 6 with missing information). Of the remainder, 3185 were eligible for inclusion in the derivation cohort and 1523 in validation cohort. Baseline characteristics are displayed in Table 1. There was no substantial intergroup difference in terms of patients' sex, age, body mass index (BMI), and left ventricular ejection fraction (LVEF). Patients with at least 3 PT-INR measurements during initial 30 days were slightly less likely to have paroxysmal AF and CHF than those with fewer than 3 INR measurements. No difference in baseline characteristics was noted between derivation and validation cohorts.

Predictive value of artificial intelligence model

ROC curve compiled for the validation cohort (Figure 4A) revealed that the AI model had a statistically higher predictive value compared with TTR with c-statistics for major bleeding and all-cause death 0.75 and 0.61, respectively (both p = 0.01 vs. TTR). A similar trend albeit nonsignificant was observed for stroke, with a c-statistic 0.70 (p = 0.08 vs. TTR). Forest plots of 95% CI for AUC of ROC curves (Figure 4B) show that the AI model performed better than random; c-statistics for major bleed, stroke, and all-cause death were 0.75 (95% CI, 0.62–0.87), 0.70 (95% CI, 0.56–0.83),

and 0.61 (95% CI, 0.54–0.67), respectively, whereas TTR was not significantly different compared with random (for same outcomes: 0.47 [95% CI, 0.32–0.61], 0.47 [95% CI, 0.31–0.64], and 0.48 [95% CI, 0.42–0.54], respectively). Table 2 shows the accuracies, sensitivities, and specificities for the best thresholds derived from the ROC curve for major bleed, stroke, and all-cause death. The model showed good predictive accuracy for major bleeding with a sensitivity 0.79 and specificity 0.78. These results were similar for the training dataset (Online Table 1, Online Figure 1A and 1B).

Survival analysis

Kaplan-Meier plots stratified by risk determined from the AI model are shown in Figure 5. The threshold of prediction score was calculated for each event to achieve the best accuracy according to the ROC curve. The best thresholds for major bleed, stroke/SE, and all-cause death were 0.27, 0.44, and 0.49, respectively. Note that these output values from our model are arbitrary numbers related to risk of future events but not actual probabilities. Patients who had a model output higher than or equal to threshold were classified as high risk and the remainder were low risk. Among 1523 patients in the validation cohort, 354 were classified high risk for major bleeding, 738 for stroke, and 560 for death. High-risk patients had higher cumulative event rates (major bleed, stroke/SE, and all-cause death) compared with low-risk patients.

The same analysis was performed for the derivation dataset. No threshold calculations were performed for the derivation dataset and the same thresholds obtained from the validation dataset were used for this analysis. The results were similar, supporting the robustness of the threshold (Online Figure 2).

Discussion

We created a new method to convert early time-series measurements of PT-INR to a long-term prediction model. The novel AI model, constructed with neural networks including 1-dimensional convolution and LSTM, garnered useful information from the raw PT-INR values and measurement dates over 30 days after VKA initiation and converted this to predict major bleeding events over the next 11 months. Although TTR is widely used to standardize VKA therapy over the long term,[14–17] its accuracy for predicting future thrombotic/bleeding events is low.[36] A previous report showed that GARFIELD-AF patients with 1-year TTR less than 65% had worse outcomes than those with greater values.[37] Within 30 days, TTR has low predictive power because early-phase PT-INR values vary greatly due to a number of influencing factors including genetics,[21,22] choice of commercial thromboplastin and coagulometer device,[38–40] and patients' lifestyles.[41] With the use of AI, we show here the presence of important information in raw PT-INR patterns over first 30 days that can predict clinical events occurring from days 31 to 365.

Multiple useful models exist to predict clinical outcomes in patients with AF.[1–3,8,42,43] However, most use single time-point data. HAS-BLED score, on the other hand, does include time-series data on PT-INR in the guise of labile PT-INR, which is expressed by TTR.[3] Our AI model using time-series PT-INR values has better predictive power than TTR for major clinical events, at least in the early phase of VKA initiation. Even with multi-dimensional data including 31 datasets our AI model output is a prediction score given as a single value. Thus, output of the new model may be included in conventional scoring models by introducing a cutoff, similarly to integration of

TTR.[18] ROC analysis in validation cohort revealed that our AI model has modest predictive power with a best c-statistic 0.78, for major bleeding. However, the model could be usefully incorporated into previous models and thereby improve their accuracy, as has been done with TTR.[18] Our prediction model could expand to automatic prediction of clinical outcomes from multi-dimensional data when incorporated into electrical patient recording systems, for example. Since our models are able to predict clinical outcomes in the early phase of treatment, they may discriminate patients who are unsuitable for VKA therapy and suggest switching them to NOACs, which are associated with lower bleeding risk compared with VKA.

Our novel AI model comprising a neural network can efficiently connect multiple time-dependent measurements to clinical outcomes to form a prediction model. Although in this study the network was used only to learn PT-INR patterns as specific target, the same structure may have the ability to convert other multi-dimensional time-dependent measurements to prediction models. Therefore, the network may provide a new means to incorporate time-dependent data in prediction models.

Output values from our AI model are related to risk of future events but not their probability. Therefore, calibration of the model with typical Hosmer-Lemeshow goodness-of-fit (GOF) test is not feasible.

Study limitations

Several limitations of this analysis should be noted. First, validation of the AI models was performed using datasets derived from the GARFIELD-AF registry. External validations of the AI model were not conducted. Thus, validity of this model beyond

GARFIELD-AF patients is unknown. On the other hand, large dissimilarities between cohorts 1–3 and 4 and 5 were noted, suggesting that our model is sufficiently robust to apply in daily clinical practice. PT-INR within 30 days may be influenced by concomitant dosing with parenteral anticoagulants. However, our model attempted to account for all influencing factors beyond the effects of VKA. We hope that other researchers will test our model's performance in external datasets.

Second, the AI model was trained only with PT-INR data, and did not include other information such as sex, age, biomarkers, concomitant drugs, or other serially measured values. Although consecutive patient data were analyzed, unrecognized confounders may exist. Many other known risk factors for adverse outcome events were not considered in our models.

Third, by selecting only patients with more than 3 PT-INR measurements within 30 days, two thirds of the entire cohort were excluded, which could introduce selection bias. Furthermore, patients do not necessarily remain stable after day 31 and our model cannot capture changes at time-points later than day 31. Future studies will examine the impact of time periods beyond 30 days in relation to AI risk prediction.

Fourth, although our results suggest the presence of crucial information within the PT-INR measurement pattern to predict patients' clinical course, the nature of that information is unknown. It might be present in the target PT-INR value, PT-INR fluctuations, PT-INR measurement frequency, or elsewhere.

Fifth. The c-statistics, sensitivity, and specificity of our models are far from perfect. Further studies to improve predictive accuracy possibly by adding other clinical characteristics and measurements are necessary.

Sixth, statistical significance was not achieved in either the derivation or validation cohort in comparison with TTR for prediction of all-cause death and stroke. This could be explained by low numbers of events limiting statistical power. Moreover, even though the number of deaths observed was not low, they could have been caused by factors not related to anticoagulation. Validation of the model in larger cohorts with higher numbers of events may demonstrate better predictive power for death and stroke. Despite these limitations, our results suggest that serially measured PT-INR values within 30 days contain information enabling us to predict serious bleeding outcomes up to 1 year. Trained AI may thus be able to detect individuals at high risk for major adverse cardiac events early in the treatment course.

Conclusions

In AF patients treated with VKA we developed new AI models to predict all-cause death, stroke, and major bleeding events occurring between months 2 and 12. The models' predictive accuracy was greatest for major bleeding, followed by all-cause mortality and stroke/SE. Our results imply that AI can capture important information to predict future outcomes from early-phase PT-INR measurements.

Acknowledgments Alex Kahney of Thrombosis Research Institute, London, UK, provided editorial support.

Funding

The GARFIELD-VTE registry is an independent academic research initiative sponsored by the Thrombosis Research Institute (London, UK) and supported by an unrestricted research grant from Bayer Pharma AG (Berlin, Germany).

Conflict of interest

S.G. (first author) has no financial competing interest to disclose. S.G. (second author) acknowledges financial support from MEXT/JSPS KAKENHI 17K19669, partly by 18H01726 and 19H03661, and from Bristol-Myers Squibb for independent research support project (33999603). S.G. (second author) acknowledges grant support from Vehicle Racing Commemorative Foundation and Nakatani Foundation for Advancement of Measuring Technologies in Biomedical Engineering. S.G. (second author) received research funding from Sanofi, Pfizer, and Ono, and a modest personal fee from Bayer. S.G. (second author) is an Associate Editor for *Circulation*, *Journal of Biorheology*, and *Archives of Medical Science* and Section Editor for *Thrombosis and Haemostasis*. K.S.P. has no financial competing interest to disclose. J.P.B. reports personal fees from Thrombosis Research Institute, during the conduct of the study. S.H. has received consulting fees and honoraria from Aspen, Bayer HealthCare, BMS/Pfizer, Daiichi-Sankyo, Portola, and Sanofi. A.P. has received consultation fees and honoraria from Bayer HealthCare, Sanofi, and Portola. F.M. is an employee of Bayer AG and a significant shareholder of Bayer shares. A.G.G.T. has received consulting fees from Bayer Health Care and honoraria from Portola and Janssen Pharmaceuticals. A.J.C. has received personal fees and institutional grants from Boehringer Ingelheim, Bayer, Daiichi Sankyo, and Pfizer/BMS. F.W.A.V. has received consulting fees and honoraria from Bayer HealthCare, Boehringer-

Ingelheim, BMS/Pfizer, and Daiichi-Sankyo. K.A.A.F. has received grants from Bayer/Janssen and AstraZeneca and consultation fees from Bayer/Janssen, Sanofi/Regeneron, and Verseon. B.J.G. is a consultant for Janssen Pharmaceuticals.

References

1. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;**137**:263–272.
2. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* 2001;**285**:2864–2870.
3. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010;**138**:1093–1100.
4. JCS Joint Working Group. Guidelines for Pharmacotherapy of Atrial Fibrillation (JCS 2013). *Circ J* 2014;**78**:1997–2021.
5. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC Jr, Ellinor PT, Ezekowitz MD, Field ME, Furie KL, Heidenreich PA, Murray KT, Shea JB, Tracy CM, Yancy CW. 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation. *Circulation* 2019:**140**;e125–e151.
6. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener HC, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, Popescu BA, Schotten U, Van Putte B, Vardas P, Agewall S, Camm J, Baron G, Esquivias W, Budts S, Carerj F, Casselman A, Coca R, De Caterina R, Deftereos S, Dobrev D, Ferro JM, Filippatos G, Fitzsimons D, Gorenek B, Guenoun M, Hohnloser SH, Kolh P, Lip GY, Manolis A, McMurray J,

Ponikowski P, Rosenhek R, Ruschitzka F, Savelieva I, Sharma S, Suwalski P, Tamargo JL, Taylor CJ, Van Gelder IC, Voors AA, Windecker S, Zamorano JL Zeppenfeld K. 2016 ESC Guidelines for the Management of Atrial Fibrillation Developed in Collaboration with EACTS. *Europace* 2016;**18**:1609–1678.

7. Lip GYH, Skjøth F, Nielsen PB, Kjældgaard JN, Larsen TB. The HAS-BLED, ATRIA, and ORBIT bleeding scores in atrial fibrillation patients using non-vitamin K antagonist oral anticoagulants. Am J Med 2018;131:574.e13–574.e27.
8. Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Hacke W, Kayani G, Oto A, Mantovani LG, Misselwitz F, Piccini JP, Turpie AGG, Verheugt FWA, Kakkar AK. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. *BMJ Open* 2017;**7**:e017157.
9. Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans Comput Biology Bioinform* 2018. doi: 10.1109/TCBB.2018.2806438.
10. Capasso R, Zurlo MC, Smith AP. Ethnicity, work-related stress and subjective reports of health by migrant workers: a multi-dimensional model. *Ethn Health* 2018;**23**:174–193.
11. Camm AJ, Accetta G, Ambrosio G, Atar D, Bassand JP, Berge E, Cools F, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Kayani G, Koretsune Y, Mantovani LG, Misselwitz F, Oh S, Turpie AG, Verheugt FW, Kakkar AK. Evolving antithrombotic treatment patterns for patients with newly diagnosed

atrial fibrillation. *Heart* 2017;**103**:307–314.

12. Kjerpeseth LJ, Ellekjaer H, Selmer R, Ariansen I, Furu K, Skovlund E. Trends in use of warfarin and direct oral anticoagulants in atrial fibrillation in Norway, 2010 to 2015. *Eur J Clin Pharmacol* 2017;**73**:1417–1425.
13. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost* 1993;**69**:236–239.
14. Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldes M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FW, Zhu J, Wallentin L. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;**365**:981–992.
15. Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, Waldo AL, Ezekowitz MD, Weitz JI, Spinar J, Ruzyllo W, Ruda M, Koretsune Y, Betcher J, Shi M, Grip LT, Patel SP, Patel I, Hanyok JJ, Mercuri M, Antman EM; Investigators EA-T. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2013;**369**:2093–2104.
16. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KA, Califf RM; Investigators RA. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;**365**:883–891.
17. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu

J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, Wallentin L. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;**361**:1139–1151.

18. Proietti M, Senoo K, Lane DA, Lip GY. Major bleeding in patients with non-valvular atrial fibrillation: impact of time in therapeutic range on contemporary bleeding risk scores. *Sci Rep* 2016;**6**:24376.
19. Williams BA, Evans MA, Honushefsky AM, Berger PB. Clinical prediction model for time in therapeutic range while on warfarin in newly diagnosed atrial fibrillation. *J Am Heart Assoc* 2017;**6**. doi: 10.1161/JAHA.117.006669.
20. Pokorney SD, Simon DN, Thomas L, Gersh BJ, Hylek EM, Piccini JP, Peterson ED. Stability of international normalized ratios in patients taking long-term warfarin therapy. *JAMA* 2016;**316**:661–663.
21. Schwarz UI, Ritchie MD, Bradford Y, Li C, Dudek SM, Frye-Anderson A, Kim RB, Roden DM, Stein CM. Genetic determinants of response to warfarin during initial anticoagulation. *N Engl J Med* 2008;**358**:999–1008.
22. International Warfarin Pharmacogenetics Consortium; Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009;**360**:753–764.
23. Wells PS, Holbrook AM, Renee Crowther N, Hirsh J. Interaction of warfarin with drugs and food. *Ann Intern Med* 1994;**121**:676–683.
24. Nathisuwan S, Dilokthornsakul P, Chaiyakunapruk N, Morarai T, Yodting T. Piriyachananusorn N. Assessing evidence of interaction between smoking and warfarin: a systematic review and meta-analysis. *Chest* 2011;**139**:1130–1139.
25. Goto S, Kimura M, Katsumata Y, Goto S, Kamatani T, Ichihara G, Ko S,

Sasaki J, Fukuda K, Sano M. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients. *PloS One* 2019;**14**:e0210103.

26. Kakkar AK, Mueller I, Bassand JP, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Hacke W, Lip GY, Mantovani LG, Verheugt FW, Jamal W, Misselwitz F, Rushton-Smith S, Turpie AG. International longitudinal registry of patients with atrial fibrillation at risk of stroke: Global Anticoagulant Registry in the FIELD (GARFIELD). *Am Heart J* 2012;**163**:13–19.e1.
27. Kakkar AK, Mueller I, Bassand JP, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Hacke W, Lip GY, Mantovani LG, Turpie AG, van Eickels M, Misselwitz F, Rushton-Smith S, Kayani G, Wilkinson P, Verheugt FW. Risk profiles and antithrombotic treatment of patients newly diagnosed with atrial fibrillation at risk of stroke: perspectives from the international, observational, prospective GARFIELD registry. *PloS One* 2013;**8**:e63479.
28. Bassand JP, Virdone S, Goldhaber SZ, Camm AJ, Fitzmaurice DA, Fox KAA, Goto S, Haas S, Hacke W, Kayani G, Mantovani LG, Misselwitz F, Pieper KS, Turpie AGG, van Eickels M, Verheugt FWA, Kakkar AK. Early risks of death, stroke/systemic embolism, and major bleeding in patients with newly diagnosed atrial fibrillation. *Circulation* 2019;**139**:787–798.
29. Bassand JP, Accetta G, Camm AJ, Cools F, Fitzmaurice DA, Fox KA, Goldhaber SZ, Goto S, Haas S, Hacke W, Kayani G, Mantovani LG, Misselwitz F, Ten Cate H, Turpie AG, Verheugt FW, Kakkar AK. Two-year outcomes of patients with newly diagnosed atrial fibrillation: results from GARFIELD-AF. *Eur Heart J* 2016;**37**:2882–2889.
30. Camm AJ, Accetta G, Ambrosio G, Atar D, Bassand JP, Berge E, Cools F,

Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Kayani G, Koretsune Y, Mantovani LG, Misselwitz F, Oh S, Turpie AG, Verheugt FW, Kakkar AK; GARFIELD-AF Investigators. Evolving antithrombotic treatment patterns for patients with newly diagnosed atrial fibrillation. *Heart* 2017;**103**:307–314.

31. Schulman S, Kearon C; Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost* 2005;**3**:692–694.
32. Taghavi Namin S, Esmaeilzadeh M, Najafi M, Brown TB, Borevitz JO. Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* 2018;**14**:66. doi: 10.1186/s13007-018-0333-4.
33. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M. Tensorflow: a system for large-scale machine learning. *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 2016:265–283.
34. Yan L, Dodier RH, Mozer M, Wolniewicz RH. Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 2003:848–855.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi: 10.1186/1471-2105-12-77.
36. Wan Y, Heneghan C, Perera R, Roberts N, Hollowell J, Glasziou P, Bankhead C, Xu Y. Anticoagulation control and prediction of adverse events in patients

with atrial fibrillation: a systematic review. *Circ Cardiovasc Qual Outcomes* 2008;**1**:84–91.

37. Haas S, Ten Cate H, Accetta G, Angchaisuksiri P, Bassand JP, Camm AJ, Corbalan R, Darius H, Fitzmaurice DA, Goldhaber SZ, Goto S, Jacobson B, Kayani G, Mantovani LG, Misselwitz F, Pieper K, Schellong SM, Stepinska J, Turpie AG, van Eickels M, Kakkar AK. Quality of vitamin K antagonist control and 1-year outcomes in patients with atrial fibrillation: a global perspective from the GARFIELD-AF Registry. *PloS One* 2016;**11**:e0164076.
38. Poller L, Ibrahim S, Keown M, Pattison A, Jespersen J; European Action on Anticoagulation. The prothrombin time/international normalized ratio (PT-INR) line: derivation of local INR with commercial thromboplastins and coagulometers—two independent studies. *J Thromb Haemost* 2011;**9**:140–148.
39. Christensen TD, Larsen TB. Precision and accuracy of point-of-care testing coagulometers for self-testing and management of oral anticoagulation therapy. *J Thromb Haemost* 2012;**10**:251–260.
40. Hemkens LG, Hilden KM, Hartschen S, Kaiser T, Didjurgeit U, Hansen R, Bender R, Sawicki PT. A randomized trial comparing INR monitoring devices in patients with anticoagulation self-management: evaluation of a novel error-grid approach. *J Thromb Thrombolysis* 2008;**26**:22–30.
41. Custódio das Dôres SM, Booth SL, Martini LA, de Carvalho Gouvêa VH, Padovani CR, de Abreu Maffei FH, Campana AO, Rupp de Paiva SA. Relationship between diet and response to warfarin: a factor analysis. *Eur J Nutr* 2007;**46**:147–154.
42. Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N, Singer DE. A new risk scheme to predict warfarin-associated hemorrhage: the

ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) Study. *J Am Coll Cardiol* 2011;**58**:395–401.

43. Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, Radford MJ. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J* 2006;**151**:713–719.

Table 1 Patients' baseline demographics and clinical characteristics*

	≥3 PT-INRs (N=4708)	0–2 PT-INRs (N=9630)	≥3 PT-INRs Subgroup	
			Derivation (N=3185)	Validation (N=1523)
Sex, n (%)				
Female	2085 (44.3)	4330 (45.0)	1420 (44.6)	665 (43.7)
Male	2623 (55.7)	5300 (55.0)	1765 (55.4)	858 (56.3)
Age at dx, years	72.1 (9.9)	70.0 (10.7)	72.2 (9.7)	72.0 (10.2)
BMI, kg/m^2	28.7 (5.9)	28.1 (5.7)	28.6 (5.7)	29.0 (6.1)
LVEF, %	53.7 (12.9)	55.7 (12.7)	53.2 (13.2)	54.7 (12.2)
Type of AF, n (%)				
New	2409 (51.2)	4087 (42.4)	1706 (53.6)	703 (46.2)
Paroxysmal	798 (16.9)	2207 (22.9)	567 (17.8)	231 (15.2)
Permanent	877 (18.6)	1514 (15.7)	487 (15.3)	390 (25.6)
Persistent	624 (13.3)	1822 (18.9)	425 (13.3)	199 (13.1)
CHF, n (%)	721 (15.3)	2149 (22.3)	466 (14.6)	255 (16.7)
CAD, n (%)	878 (18.6)	1896 (19.7)	511 (16.0)	367 (24.1)
ACS	461 (9.8)	872 (9.1)	292 (9.2)	169 (11.1)
CHA_2DS_2-VASc	3.4 (1.5)	3.3 (1.5)	34 (1.5)	33 (1.4)
HAS-BLED	1.4 (0.9)	1.4 (0.9)	15 (0.9)	14 (0.9)

*Values are mean (SD) unless specified otherwise.

ACS, acute coronary syndromes; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; LVEF, left ventricular ejection fraction.

Table 2 Best predictive accuracies and corresponding sensitivities and specificities (95% CIs) for validation cohort

AI	Accuracy	Sensitivity	Specificity
Major bleed	0.78 (0.40–0.92)	0.79 (0.50–1.00)	0.78 (0.39–0.93)
Stroke	0.53 (0.24–0.98)	0.85 (0.31–1.00)	0.53 (0.23–0.99)
All-cause death	0.64 (0.51–0.69)	0.63 (0.50–0.76)	0.65 (0.50–0.70)

Figure legends

Figure 1 Structure of the neural network and input data for the model. Schematic illustration of the neural network model (A). Vector of 30 PT-INR measurements for 4708 patients used as input to the model (B). The element at n^{th} dimension of the vector holds the PT-INR value measured on day n. Un-measured data-points were filled with 0. These pairs of PT-INR vectors labeled with outcome for each patient were used for model training (C). LSTM, long short-term memory; 1D convolution, one-dimensional convolution; INR, prothrombin time international normalized ratio.

Figure 2 Model training process. Schematic illustration of the training process. Datasets are indicated with yellow, processes indicated with blue, and models derived in each epoch indicated with light red. The derivation cohort was further split into training dataset and testing dataset. Training was performed by mini-batch method using 455 as batch size. For each epoch, the derived model was compared with the best model derived with previous epochs using the testing datasets. Training was performed for 500 epochs and the best model using the testing dataset selected.

Figure 3 Patient selection. Schematic illustration of patient selection process. VKA, vitamin K antagonist; PT-INR, prothrombin time international normalized ratio.

Figure 4 ROC analysis of the AI model. Comparison of ROC curves compiled from AI model and TTR (A). Comparisons were performed using stratified bootstrap method with 2000 bootstrap rounds. Forest plots of AUC of the ROC curve for each outcome (B). The 95% CIs were calculated by stratified bootstrap method with 2000

bootstrap rounds. ROC, receiver operating characteristic; AUC, area under curve; CI, confidence interval.

Figure 5 Cumulative event rates according to risk stratified by AI model. Kaplan-Meier curves stratified with risk according to AI model for major bleed (A), stroke/SE (B), and all-cause death (C). The threshold for risk stratification was made by that which gave the best accuracy in the validation cohort for each outcome according to the ROC curve. The p-values were calculated by log-rank test. ROC, receiver operating characteristic; AUC, area under curve.

Figure 1

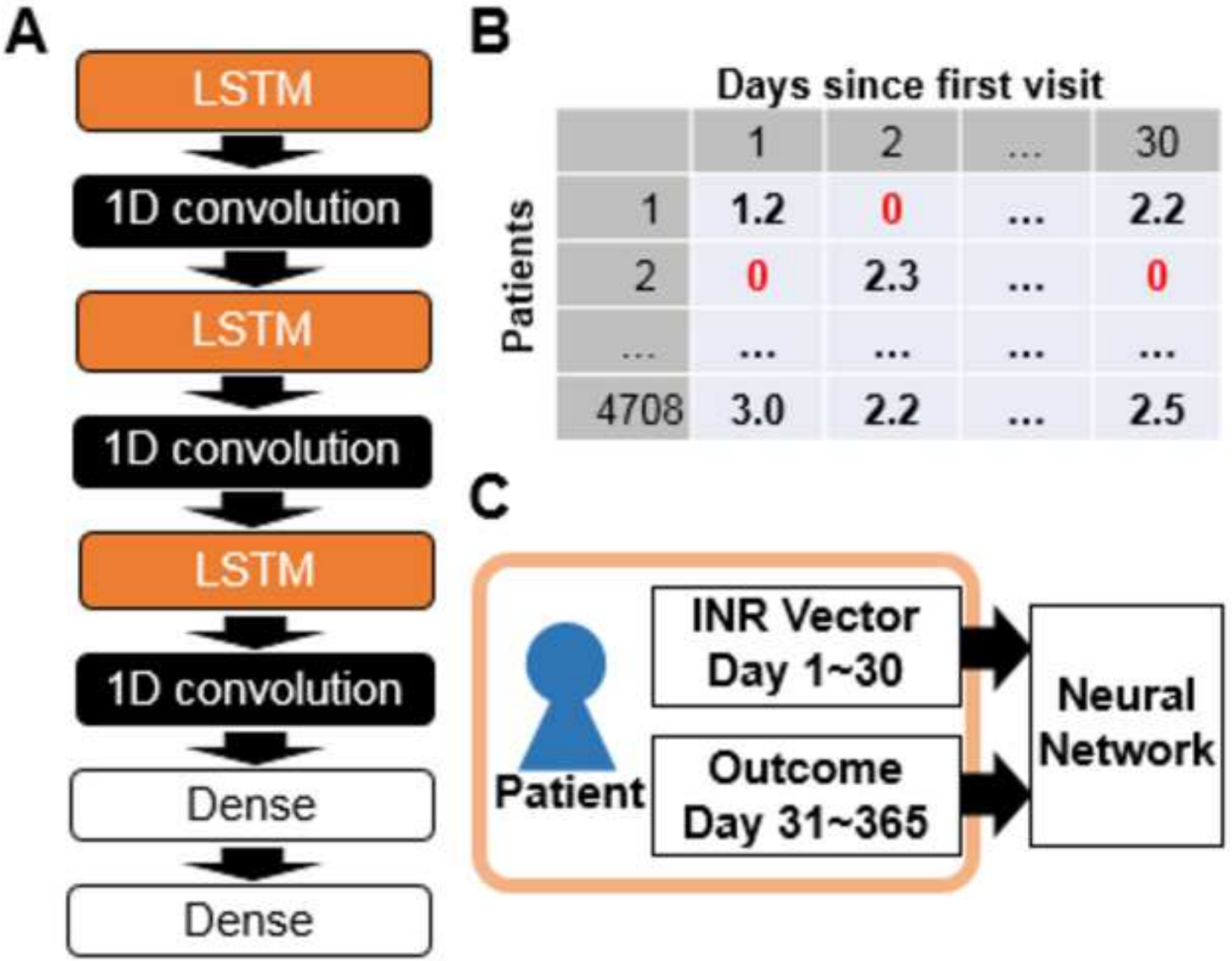

Figure 2

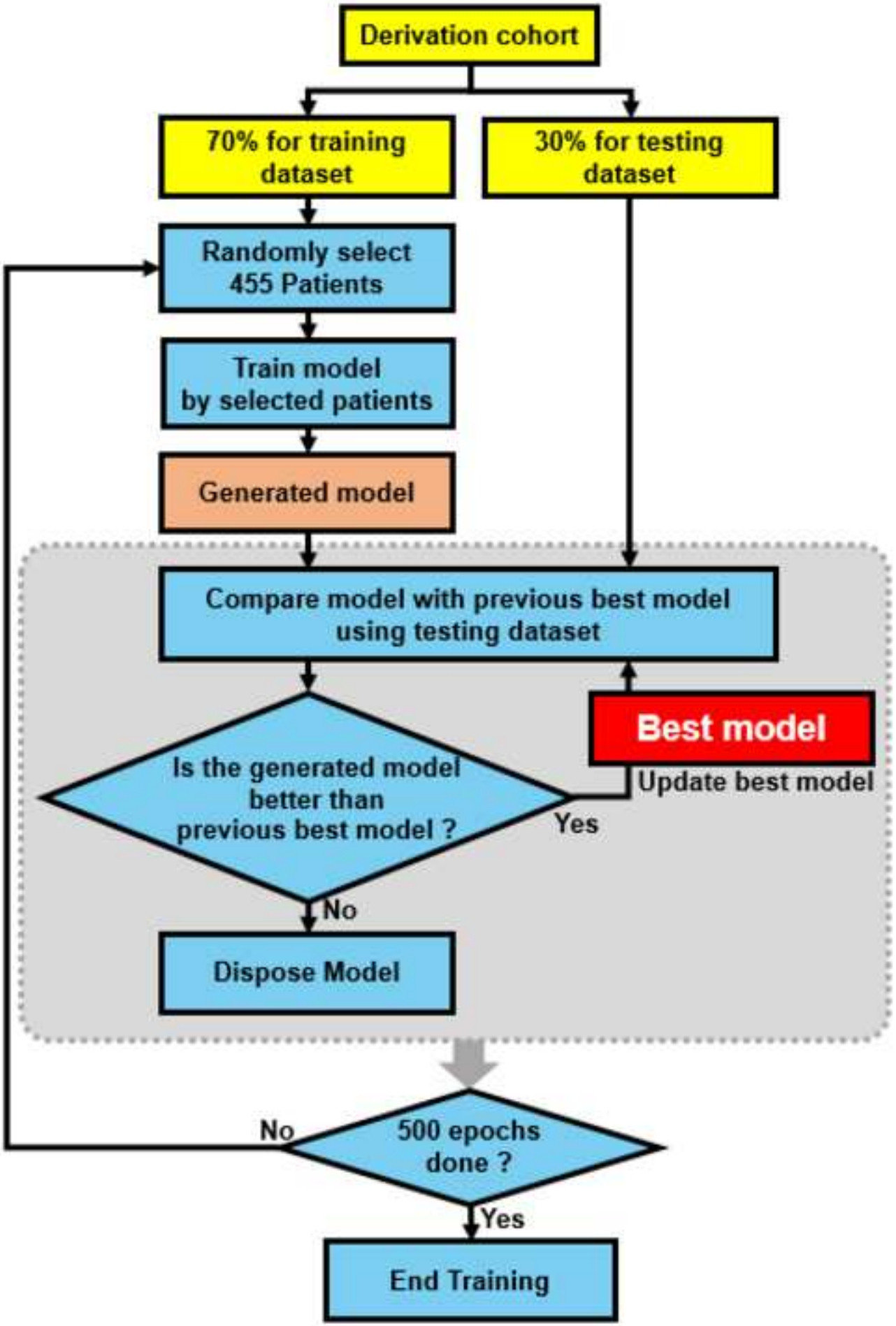

Figure 3

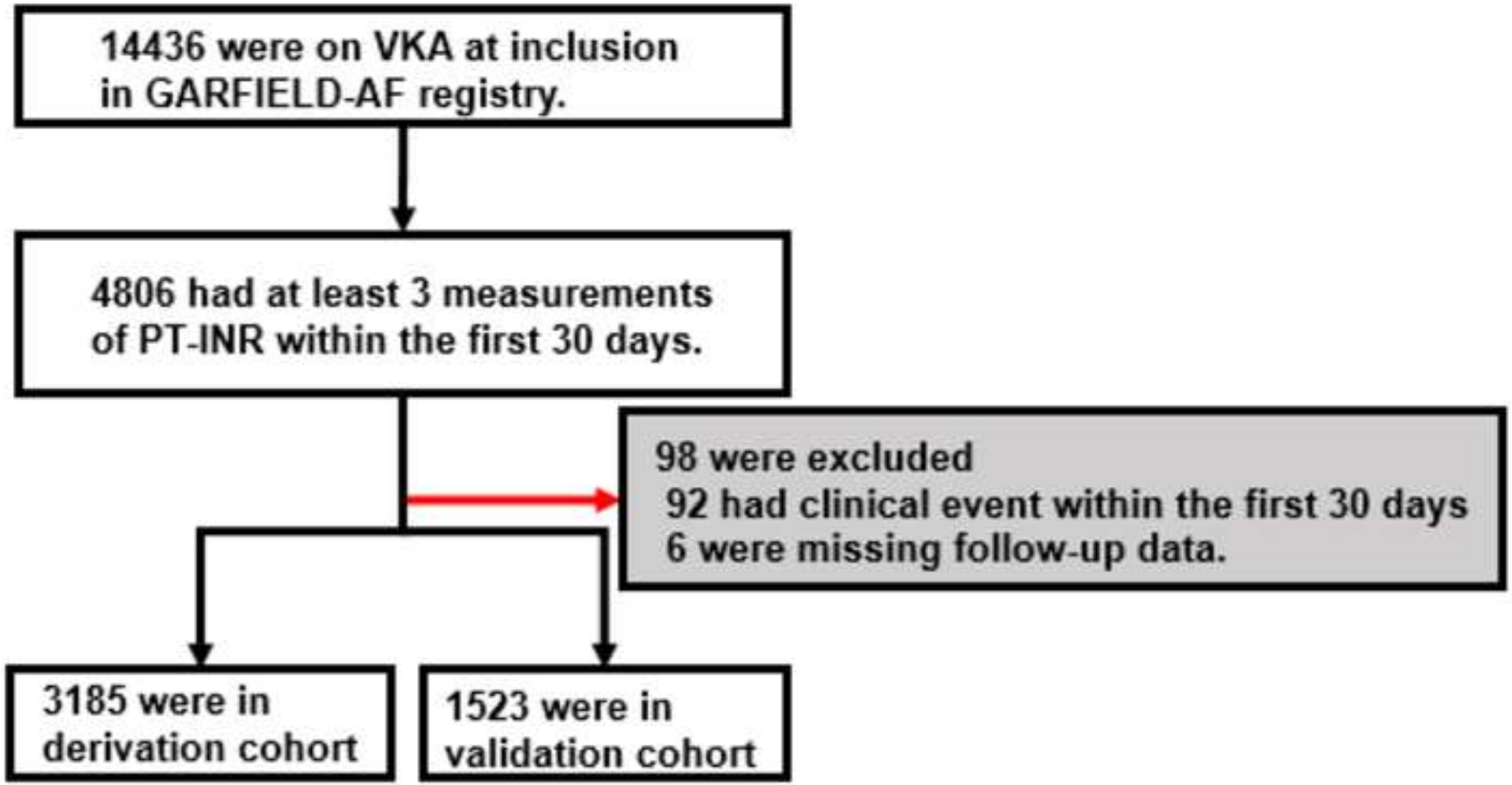

Figure 4

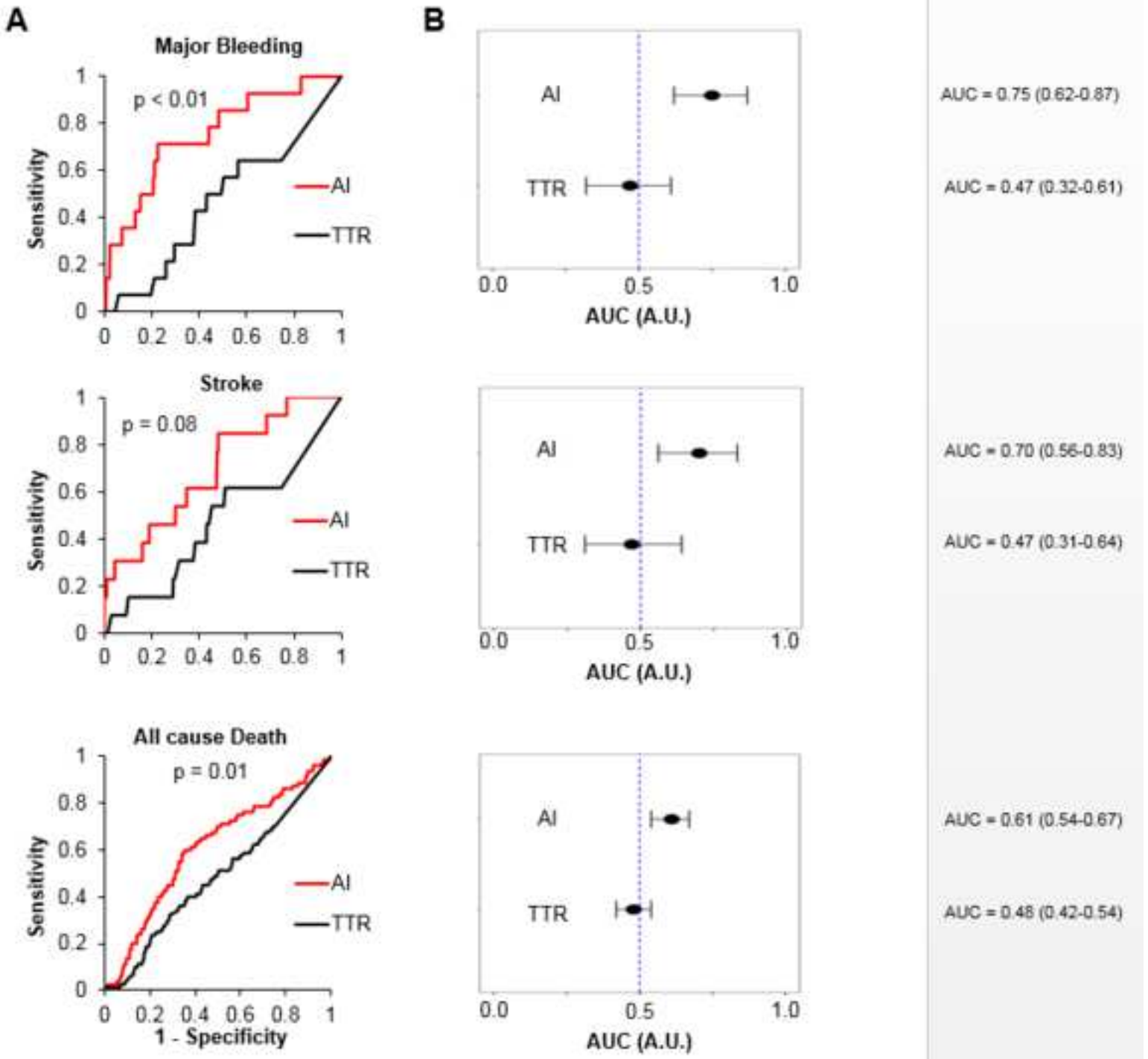

Figure 5