

Real-time prospective (shadow mode) validation of an AI-based clinical decision support system for predicting 3-month functional outcome in acute stroke: the VALIDATE study protocol

Marta Rubiera, MD, PhD; Martin Bendszus*, MD, PhD; Ronen R. Leker* MD, PhD; Adam Hilbert, MSc; Ingo Werren, PhD; Luis M. Lopez-Ramos, PhD; Mercedes Ayesta, RN; Thi Nguyet Que Nguyen, MSc; Sussane Bonekamp, PhD; Victoria Sala, RN; Hamza Jubran MD, Claudia Meza, RN; Fatma Shalabi MD, Yoel Schwartzmann MD, David Cano, DS; Malte von Tottleben, MSc; John Kelleher, PhD and Dietmar Frey, MD, PhD for the VALIDATE investigators.

***Equal contribution**

AFFILIATIONS:

Hospital Universitari Vall d'Hebron, Stroke Unit, Barcelona, Spain (MR); Vall d'Hebron Institut de Recerca, Stroke Research Group, Barcelona, Spain (MR, VS, CM)

University of Heidelberg, Department of Neuroradiology, Heidelberg, Germany (MB, SB)

Hebrew University Hadassah Medical School, Department of Neurology, Yerushalayim, Israel (RRL, HJ, FS, YS)

Charité Universitätsmedizin, Charité Lab for Artificial Intelligence in Medicine (CLAIM), Berlin, Germany (AH, DF)

IBM iX GmbH, Digital Health Department, Berlin, Germany (IW)

Simula Metropolitan Center for Digital Engineering, Oslo, Norway; Rey Juan Carlos University – Fuenlabrada School of Engineering, Fuenlabrada, Spain (LML-R)

NORA Health, Sant Cugat, Spain (MA, DC)

Technological University Dublin, Dublin, Ireland (TNQN)

Empirica Gesellschaft für Kommunikations- und Technologieforschung mbH, Digital Health, Bonn, Germany (MvT)

Trinity College, ADAPT Research Centre, School of Computer Science and Statistics, Dublin, Ireland (JK)

Department of Neurosurgery, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Germany (DF)

Corresponding author:

Marta Rubiera, MD, PhD

Stroke Neurologist.

Stroke Unit - Neurology Department. Hospital Universitari Vall d’Hebron.

Passeig Vall d’Hebron 119-129. 08035-Barcelona. Spain

Telephone: 0034 647390477

Key words: Artificial intelligence, outcome prediction, acute stroke

Word count: Abstract 299, Manuscript body 2096, total 3647

Tables: 1; **Figures:** 2; **Supplemental material:** TRIPOD-AI checklist, Statistical Analysis Plan, 2 tables, 1 Figure

Abstract:

Introduction

Despite the proven benefits of reperfusion therapies in acute ischemic stroke, treatment decisions in the hyperacute phase remain complex and are rarely supported by individualized outcome predictions. Artificial intelligence (AI)-based clinical decision support systems (CDSS) offer potential real-time prognostic estimates, but prospective evidence of their feasibility and performance in routine clinical workflows is limited. Our aim is to prospectively evaluate real-time feasibility, usability, and predictive performance of an AI-based CDSS (VALIDATE-CDSS) for individualized outcome prediction in acute stroke care.

Methods and analysis

Prospective, multicenter, observational study enrolling consecutive patients with acute ischemic stroke presenting to three tertiary stroke centers. Clinical management will follow standard practice at the discretion of treating physicians. In parallel, a dedicated researcher will collect patient data in real time and input them into the VALIDATE-CDSS using a mobile application, operating in shadow mode without influencing clinical decisions. The system will generate individualized predictions of 3-month functional outcome (modified Rankin Scale) for four treatment strategies (intravenous thrombolysis, endovascular thrombectomy, combined therapy, or no reperfusion) at three sequential time points: baseline clinical data, non-contrast CT, and CT angiography.

The primary outcome is the real-world feasibility and usability of the VALIDATE-CDSS in the hyperacute stroke workflow. Secondary outcomes include predictive performance, agreement between model-suggested and actual treatments, incremental value with increasing data availability, and assessment of potential bias across predefined subgroups.

This study will provide prospective real-world evidence on the implementation and clinical potential of AI-based decision support for personalized treatment selection in acute ischemic stroke

Ethics and dissemination

Patient enrollment began after approval from the ethics committees of all participating centers. Results will be disseminated through peer-reviewed open-access journals and conference presentations. Following open science principles, anonymized data and metadata will be made publicly available in the Zenodo repository upon study completion.

Trial registration: ClinicalTrials.gov (NCT05622539).

Strengths and limitations of this study

- First study to assess the feasibility of integrating an outcome-predictor CDSS into real-life hyperacute stroke workflows, addressing a critical gap between AI model development and clinical implementation
- Multicenter, prospective observational time-motion shadow-mode design, which minimizes interference with standard care while capturing real-world operational data

- Validation of a locked AI model developed from independent retrospective multicenter datasets across different populations, reducing the risk of overfitting to local case-mix
- Real-time data acquisition in the emergency department poses a significant operational challenge, with potential for missing or delayed inputs that may affect model performance in practice
- Risk of bias cannot be excluded, including spectrum bias from non-anticipated subgroups, temporal drift in clinical practice or patient populations, and centre-level variation in workflow and data quality

Introduction

Artificial intelligence (AI)–based prognostic tools and clinical decision support systems (CDSS) can rapidly analyze large volumes of patient data to predict disease outcomes, supporting clinical decision-making. These tools enable more accurate and individualized treatment stratification, with the potential to improve patient outcomes and optimize healthcare resource use (1).

In acute ischemic stroke, the population-level benefits of intravenous thrombolysis (IVT) and mechanical endovascular thrombectomy (EVT) are well established (2). However, individual outcomes vary widely, and a considerable proportion of treated patients still experience poor recovery (3). Predicting outcomes in the hyperacute stroke setting is particularly challenging due to time constraints and the complex interaction of multiple prognostic factors (4). AI methods are well suited to this context, as they can identify complex, non-linear relationships across numerous variables and patient subgroups, and can be adapted for use across different levels of care (5).

VALIDATE, a Horizon Europe–funded project (No. 101057263) aims to test the feasibility of such AI integration into the acute stroke workflow. AI developers in the project together with Stroke Action For Europe patients’ association (SAFE) and clinical partners designed and created a machine learning based model to predict the modified Rankin Scale (mRS) (6) 3 months after stroke. The AI model has been fully developed and internally validated prior to this study using independent retrospective multicenter datasets from different populations, including retrospective data from the three comprehensive stroke centers (CSCs) participating in the present study (7). The model will remain locked during the study, with no further training or recalibration performed. For the present study, the model has been integrated into a CDSS combining structured data collection and real-time prediction.

We hypothesize that an AI-based CDSS providing individualized long-term outcome predictions can be feasibly integrated into real-world hyperacute stroke care workflows, and that it will demonstrate clinically meaningful predictive accuracy across the different hyperacute treatment options.

Methods and analysis

Study design

VALIDATE is a multicenter, prospective, phase 2 observational time-motion shadow-mode study conducted in three high-volume CSCs: two in Europe (Germany and Spain) and one in Israel. Consecutive acute ischemic stroke patients will be recruited over a 1.5-year period.

Patient and Public Involvement Statement

The VALIDATE consortium includes the Stroke Action For Europe patients' association (SAFE) as a full partner. SAFE has been involved since the earliest stages of project design, contributing particularly to the definition of outcomes that are meaningful to patients within a value-based healthcare framework. SAFE also plays a key role in dissemination, ensuring that results reach not only the clinical community but also healthcare institutions, patients, and their families.

Patient population and study intervention

All patients admitted with acute ischemic stroke (<24 hours from symptom onset or stroke of unknown onset) will be screened for inclusion. Clinical management and treatment decisions will be performed according to local protocols and at the discretion of the treating stroke physicians.

After stroke code activation, a stroke-trained shadow researcher (SR) will accompany the treating team in the emergency department to enroll eligible patients, operating in a prospective time-motion “shadow mode” design. The SR will not display predictions or interact with treating clinicians, ensuring that the AI system does not influence clinical decision-making at any stage. Inclusion and exclusion criteria are detailed in Table 1.

Written informed consent will be obtained from the patient, a legal representative, a family member, or an independent physician when the patient is unable to consent. Re-consent will be sought during hospitalization if the patient later regains decision-making capacity.

For each enrolled patient, relevant clinical and imaging data will be entered in real time into a dedicated mobile application (VALIDATE app, see Figure 1).

Variables required for the AI prediction will be manually entered by the SR during the hyperacute phase without interfering with routine care. The VALIDATE-CDSS generates probabilistic predictions of the mRS at 3 months, even when some variables are missing. Predictions will be generated at three predefined time points along the hyperacute workflow (see Figure 2). After non-contrast computerized tomography (NCCT), patients with intracranial hemorrhage or non-ischemic pathology will be excluded from the study.

Predictions will be manually triggered by the SR via the app. Data will be transmitted through a closed local wireless network to a dedicated workstation hosting the AI model, which will act as a backend server, and will enable triggering predictions and collecting data (Figure S1 from Online Supplemental). Prediction results will be returned and visualized in the app. This isolated local configuration complies with data protection requirements mandated by the ethics committees. In addition, the time-

motion design will allow detailed characterization of workflow integration and operational impact of the AI system in the hyperacute setting.

The app displays outcome probabilities across the full mRS spectrum (scores 0–6) and as favorable/unfavorable outcomes using an adjustable threshold, stratified by treatment strategy: IVT, EVT, combined IVT+EVT, or no reperfusion treatment. Model confidence and visual representations of the contribution of different clinical variables to the predicted probability of treatment success (“data impact”) are also provided for individual predictions.

For each prediction, a structured report including predicted mRS probabilities, input variables, confidence metrics, and timestamps will be automatically generated and stored locally in the backend host workstation in CSV format.

Treating physicians will remain blind to all predictions. Patients will continue standard in-hospital care and follow-up. The actual mRS score at 3 months will be assessed by certified personnel, blinded to predictions, via telephone or in-person using standardized procedures. Additional outcomes, including patient-reported outcome measures (PROMs) will be collected from electronic medical records, telephone interviews, or through the NORA (8) digital platform, which enables patient education, communication, and survey-based PROM collection. A complete list of data collected for the study can be found in Table S.1 (Online Supplemental).

Study endpoints

The primary objective is to assess the real-world feasibility and usability of real-time AI-based outcome prediction within the hyperacute stroke workflow.

Secondary objectives include evaluating the accuracy of AI predictions for 3-month functional outcome measured by the mRS across the full scale, dichotomized (mRS 0–2

vs. 3–6), and trichotomized (0–2, 3–4, 5–6) distributions. Additional analyses will assess concordance between predicted optimal treatment and actual treatment administered, correlation with PROMs, improvement in model performance with increasing data availability, and potential biases across predefined clinical and demographic subgroups.

Data management and security

Given the observational design, no data safety monitoring board is required. A clinical research manager will oversee study conduct, with regular meetings and centralized monitoring through the NORA platform.

An electronic case report form (eCRF) integrated into NORA will store all study data. Acute-phase data collected by SR, outcomes, and key performance indicators will be consolidated into center-specific databases hosted on local hospital servers. After study completion, data will be anonymized locally and merged into a centralized dataset stored on secure servers at the Vall d’Hebron Institute of Research. All data handling complies with the European General Data Protection Regulation and relevant national regulations.

Statistical analysis

Feasibility will be assessed by system penetration rate (target $\geq 70\%$ during working hours), technical failures ($< 10\%$ of technical failure during hyperacute phase), and door-to-prediction time. Usability will be evaluated using the System Usability Scale (target ≥ 68) and targeted surveys (see online supplemental Table S.2).

For the secondary outcomes, Cohen’s Kappa statistics will be used to measure the level of agreement between the treatment option with the better mRS predicted by the ML algorithms and real treatment received by the patient. Cohen’s Kappa will also be used

to compare the outcome predicted by the VALIDATE-CDSS for the treatment received by the patient and his/her actual outcome.

Predictive performance will be primarily assessed using discrimination (AUROC), calibration, and agreement metrics, with additional classification metrics reported as secondary measures. Feature impact analysis will also be conducted. Following EU open science recommendations, anonymized data and metadata will be deposited in Zenodo after study completion. A complete statistical analysis plan can be found in the Supplemental Online Material.

Sample size estimation

Sample size for this feasibility study was not determined by formal power calculation, as the primary aim is not to test a hypothesis but to estimate key feasibility parameters with sufficient precision to inform the design of a subsequent definitive trial. No formal hypothesis testing is planned for the primary endpoint.

A sample of 200 patients (approximately 67 per site over a recruitment period of 18 months) will allow estimation of the system penetration rate, the primary feasibility endpoint, with a 95% confidence interval of ± 7 points around the expected target of 70% (based on a normal approximation for proportions). This precision is considered adequate to determine whether the pre-specified feasibility threshold has been met and to detect site-level variation in adoption. This approach is consistent with established guidance on the design and reporting of feasibility and pilot studies, which recommend that sample sizes be justified on the basis of estimation precision rather than statistical power.

Ethics and Dissemination

The study is registered at ClinicalTrials.gov (NCT05622539), and patient enrollment started in July 2024, after approval from the Ethics Committees of the three clinical sites (Vall d'Hebron Institut de Recerca, PR(AG)432-2023; Universität Heidelberg Ethikkommission der Med. Fakultät, S-687/2023 and Helsinki Committee (Institutional Review Board) of Hadassah Medical Center-Ein Kerem, HMO-0529-22). This protocol was developed in accordance with the TRIPOD (9,10) guidelines for multivariable prediction models. All participating centers adhere to current European and international guidelines for acute stroke management (11). The protocol is publicly available in Zenodo (<https://zenodo.org/records/17367485>).

Results will be disseminated through peer-reviewed open-access journals and conference presentations. In line with open science principles, anonymized data and metadata will be made publicly available in the Zenodo repository upon study completion. The analytical code will be available after the future publication of the algorithms.

Discussion

Predicting long-term outcomes for patients with acute ischemic stroke at emergency department admission remains highly challenging due to the interplay of multiple clinical factors, such as age, sex, pre-existing disability, stroke severity, time from symptom onset, and the variable effects of hyperacute treatments (12). Although randomized trials have demonstrated the efficacy of reperfusion therapies in selected populations, their strict inclusion criteria limit applicability in real-world settings (4). Consequently, clinical decision-making is often complex, especially for less experienced physicians or for patients who fall outside trial parameters. This complexity is intensified by the extreme time pressure of hyperacute care, leading to heterogeneous outcomes (13).

In this scenario, CDSS such as VALIDATE may assist clinicians by providing individualized outcome predictions under different treatment strategies. These tools can support real-time decisions and improve communication with patients and families by clarifying potential benefits and risks. Beyond acute care, CDSS may also enhance education, quality assurance, and continuous improvement of stroke services (14).

Previous predictive efforts have relied mainly on retrospective datasets and traditional statistical or AI-based models, typically offering dichotomized mRS predictions with accuracies between 0.70 and 0.80. Performance improves when incorporating treatment or early post-treatment data, but most models are developed in homogeneous cohorts and tested outside real-time workflows (15). In contrast, VALIDATE-CDSS was trained on large, multinational, multiracial datasets (7) and will be evaluated prospectively during hyperacute stroke care.

This study is the first to assess the feasibility of integrating a CDSS into real-life hyperacute workflows using an observational time-motion design, with a shadowing researcher accompanying the clinical team without interaction. This approach allows precise assessment of timing, integration, and usability while preserving the observational nature of the study by blinding physicians to predictions. Although shadowing methodologies are well established in behavioral sciences, their application to time-critical acute clinical workflows has been limited.

The target sample of 200 patients across three international centres over a recruitment period of 18 months reflects the expected volume of eligible acute stroke patients presenting during working hours at each participating site, based on historical admission rates. As this is a feasibility study, no formal power calculation was performed. The sample size is sufficient to assess system penetration, usability, and operational performance across diverse clinical settings, and to identify site-specific barriers to

implementation that would inform the protocol of a future adequately powered trial (16).

A major challenge is real-time data acquisition, as some variables require bedside assessment and manual entry. VALIDATE-CDSS was therefore designed for rapid data input, local AI execution, and immediate result visualization without delaying treatment. The study will help identify potential workflow challenges introduced by AI-based CDSS.

Summary

VALIDATE will evaluate the feasibility of a CDSS that provides individualized long-term outcome predictions for hyperacute ischemic stroke patients across different reperfusion strategies. If feasibility is demonstrated, these results will support the design of a randomized clinical trial comparing standard clinical decision-making with and without AI-based decision support.

References

1. Gonzalez-Smith J, Shen H, Singletary E, Silcox C. How health systems decide to use artificial intelligence for clinical decision support. *NEJM Catal.* 2022;3(4). doi:10.1056/CAT.21.0416
2. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2019;50(12):e344-e418. doi:10.1161/STR.0000000000000211
3. Goyal M, Menon BK, van Zwam WH, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* 2016;387(10029):1723-1731. doi:10.1016/S0140-6736(16)00163-X
4. Pantanowitz L, Pearce T, Abukhiran I, et al. Nongenerative artificial intelligence in medicine: advancements and applications in supervised and unsupervised machine learning. *Mod Pathol.* 2025;38(3):100680. doi:10.1016/j.modpat.2024.100680
5. Mouridsen K, Thurner P, Zaharchuk G. Artificial intelligence applications in stroke. *Stroke.* 2020;51(8):2573-2579. doi:10.1161/STROKEAHA.119.027479

6. Huybrechts KF, Caro JJ. The Barthel Index and modified Rankin Scale as prognostic tools for long-term outcomes after stroke: a qualitative review of the literature. *Curr Med Res Opin.* 2007;23(7):1627-1636. doi:10.1185/030079907x210444
7. Hilbert A, Rieger J, Aydin OU, Nijenhuis FT, Rubiera M, Bonekamp S, Leker RR, Tanioka S, Walsum T, Majoie C, Zwam W, Payabvash S, Tourdias T, Nolte C, Frey D. VALIDATE: Large Scale Retrospective Validation of AI-Based Outcome Prediction in Ischemic Stroke. *Eur Stroke J.* 2025; 10 (2_suppl, 1): 289-290.
8. Requena M, Montiel E, Baladas M, et al. Farmalarm. *Stroke.* 2019;50(7):1819-1824. doi:10.1161/STROKEAHA.118.024355
9. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation.* 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508
10. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73. doi:10.7326/M14-0698
11. Turc G, Tsivgoulis G, Audebert HJ, et al. European Stroke Organisation - European Society for Minimally Invasive Neurological Therapy expedited recommendation on indication for intravenous thrombolysis before mechanical thrombectomy in patients with acute ischaemic stroke and anterior circulation large vessel occlusion. *Eur Stroke J.* 2022;7(1):I-XXVI. doi:10.1177/23969873221076968
12. Goyal M, Ospel JM, Kappelhof M, Ganesh A. Challenges of outcome prediction for acute stroke treatment decisions. *Stroke.* 2021;52(5):1921-1928.

doi:10.1161/STROKEAHA.120.033785

13. Ramos LA, Kappelhof M, van Os HJA, et al. Predicting poor outcome before endovascular treatment in patients with acute ischemic stroke. *Front Neurol.* 2020;11:580957. doi:10.3389/fneur.2020.580957
14. Chen Z, Liang N, Zhang H, et al. Harnessing the power of clinical decision support systems: challenges and opportunities. *Open Heart.* 2023;10(2). doi:10.1136/openhrt-2023-002432
15. Ramos LA, van Os H, Hilbert A, et al. Combination of radiological and clinical baseline data for outcome prediction of patients with an acute ischemic stroke. *Front Neurol.* 2022;13:809343. doi:10.3389/fneur.2022.809343
16. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, Lancaster GA; on behalf of the PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ.* 2016;355:i5239. <https://doi.org/10.1136/bmj.i5239>

Author Contributions

Marta Rubiera: Conceptualization, Methodology, Supervision, Writing – Original Draft;
Martin Bendszus: Conceptualization, Supervision, Writing – Review&Editing; Ronen
R. Leker: Conceptualization, Supervision, Writing – Review&Editing; Adam Hilbert:
Conceptualization, Funding Acquisition. Software, Validation, Writing –
Review&Editing; Ingo Werren: Conceptualization, Software, Writing –
Review&Editing; Luis M. Lopez-Ramos: Conceptualization, Software, Validation,
Writing – Review&Editing; Mercedes Ayesta: Data Curation, Writing –
Review&Editing; Thi Nguyet Que Nguyen: Software, Validation, Writing –
Review&Editing; Sussane Bonekamp: Investigation, Writing – Review&Editing;
Victoria Sala: Investigation, Writing – Review&Editing; Hamza Jubran: Investigation,
Writing – Review&Editing; Claudia Meza: Investigation, Writing – Review&Editing;
Fatma Shalabi: Investigation, Writing – Review&Editing; Yoel Schwartzmann:
Investigation, Writing – Review&Editing; David Cano: Data Curation, Writing –
Review&Editing; Malte von Tottleben: Project Administration, Writing –
Review&Editing; John D. Kelleher: Conceptualization, Methodology, Writing –
Review&Editing; Dietmar Frey: Conceptualization, Methodology, Supervision,
Funding Acquisition.

All authors read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This work was supported by the European Commissions' Horizon Europe Programme grant number 101057263.

Disclosures

Dr Rubiera reported compensation from Bayer for data and safety monitoring services outside the present work. Dr. Lecker received speaker honoraria from IschemaView, Boehringer Ingelheim, Pfizer, Jansen, Biogen, Medtronic and Abbott and advisory board honoraria from Jansen, all outside the submitted work. The remaining authors report no disclosures related to the present work.

Acknowledgements

We thank all the VALIDATE consortium investigators and collaborators

Table 1: Inclusion and exclusion criteria of the prospective study.

INCLUSION CRITERIA	EXCLUSION CRITERIA
<ul style="list-style-type: none">- Age \geq 18 years- Acute ischemic stroke $<$24 hours from symptoms onset or unknown onset- NIHSS $>$1- Informed consent	<ul style="list-style-type: none">- Known acute intracranial hemorrhage (ICH)- Patients transferred from other facilities without repeated CT- Serious, advanced, or terminal illness with anticipated life expectancy of less than 3 months- Unlikely to be available for 90-day follow-up (e.g. no fixed home address, no telephone,..)

Figure 1: VALIDATE app: input data and prediction visualization.

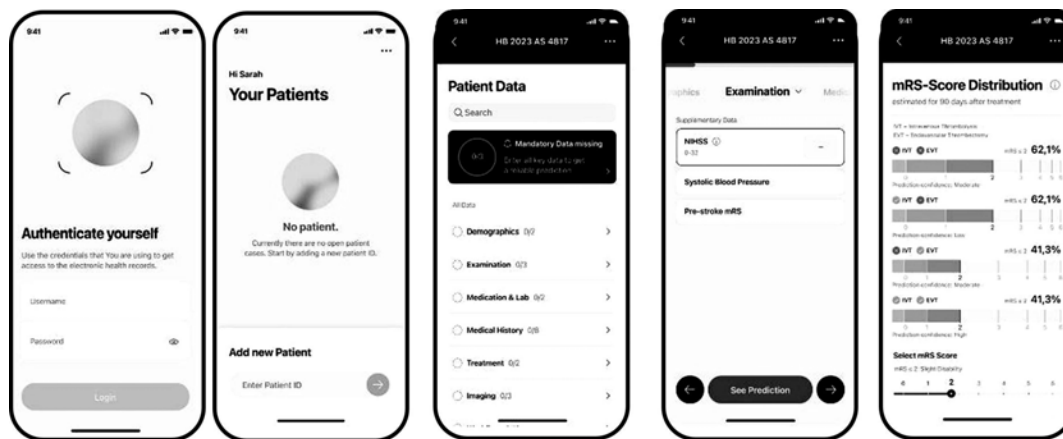
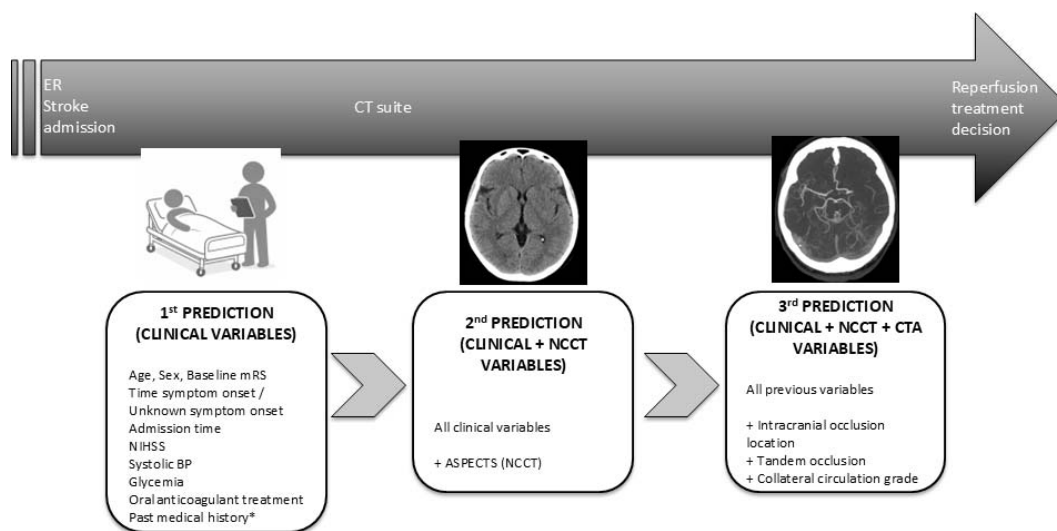


Figure 2: Workflow for sequential outcome prediction in the VALIDATE protocol



* Past medical history includes: hypertensive, Diabetes Mellitus, hypercholesterolemia, smoking, atrial fibrillation, ischemic heart disease, valvular heart disease, previous stroke

mRS: modified Rankin Scale; NIHSS: National Institute of Health Stroke Scale; BP: blood pressure; NCCT: non-contrast computerized tomography; ASPECTS: Alberta Stroke Program CT score; CTA: CT angiography