

2023

## Comparing Poor and Favorable Outcome Prediction With Machine Learning After Mechanical Thrombectomy in Acute Ischemic Stroke

Matthias A. Mutke

*Heidelberg University Hospital, Heidelberg, Germany*

Vince I. Madai

*Charite Universitatsmedizin Berlin, Berlin, Germany*

Adam Hilbert

*Charite Universitatsmedizin Berlin, Berlin, Germany*

*See next page for additional authors*

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomart>



Part of the [Computer Engineering Commons](#)

### Recommended Citation

Mutke, Matthias A.; Madai, Vince I.; Hilbert, Adam; Zihni, Esra; Potreck, Arne; Weyland, Charlotte S.; Mohlenbruch, Markus A.; Heiland, Sabine; Ringleb, Peter A.; Nagel, Simon; Beendszus, Martin; and Frey, Dietmar, "Comparing Poor and Favorable Outcome Prediction With Machine Learning After Mechanical Thrombectomy in Acute Ischemic Stroke" (2023). *Articles*. 189.

<https://arrow.tudublin.ie/scschcomart/189>

This Article is brought to you for free and open access by the School of Computer Science at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [vera.kilshaw@tudublin.ie](mailto:vera.kilshaw@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-Share Alike 4.0 International License](#).

Funder: ai4medicine; European Commission Horizon2020 PRECISE4Q No. 777107.

---

## Authors

Matthias A. Mutke, Vince I. Madai, Adam Hilbert, Esra Zihni, Arne Potreck, Charlotte S. Weyland, Markus A. Mohlenbruch, Sabine Heiland, Peter A. Ringleb, Simon Nagel, Martin Beendszus, and Dietmar Frey



# Comparing Poor and Favorable Outcome Prediction With Machine Learning After Mechanical Thrombectomy in Acute Ischemic Stroke

Matthias A. Mutke<sup>1\*</sup>, Vince I. Madaj<sup>2,3,4</sup>, Adam Hilbert<sup>2</sup>, Esra Zihni<sup>2,5</sup>, Arne Potreck<sup>1</sup>, Charlotte S. Weyland<sup>1</sup>, Markus A. Möhlenbruch<sup>1</sup>, Sabine Heiland<sup>1</sup>, Peter A. Ringleb<sup>6</sup>, Simon Nagel<sup>6</sup>, Martin Bendszus<sup>1</sup> and Dietmar Frey<sup>2</sup>

<sup>1</sup> Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany, <sup>2</sup> Charité Lab for Artificial Intelligence in Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany, <sup>3</sup> QUEST (Quality, Ethics, Open Science, Translation) Center for Responsible Research at Berlin Institute of Health, Charité Universitätsmedizin Berlin, Berlin, Germany, <sup>4</sup> School of Computing and Digital Technology, Faculty of Computing, Engineering and the Built Environment, Birmingham City University, Birmingham, United Kingdom, <sup>5</sup> School of Computing, Technological University Dublin, Dublin, Ireland, <sup>6</sup> Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany

## OPEN ACCESS

### Edited by:

Tae-Hee Cho,  
Hospices Civils de Lyon, France

### Reviewed by:

Bum Joon Kim,  
University of Ulsan, South Korea  
Carole Frindel,  
Université de Lyon, France

### \*Correspondence:

Matthias A. Mutke  
matthias.mutke@  
med.uni-heidelberg.de

### Specialty section:

This article was submitted to  
Stroke,  
a section of the journal  
Frontiers in Neurology

Received: 07 July 2021

Accepted: 28 March 2022

Published: 27 May 2022

### Citation:

Mutke MA, Madaj VI, Hilbert A, Zihni E, Potreck A, Weyland CS, Möhlenbruch MA, Heiland S, Ringleb PA, Nagel S, Bendszus M and Frey D (2022) Comparing Poor and Favorable Outcome Prediction With Machine Learning After Mechanical Thrombectomy in Acute Ischemic Stroke. *Front. Neurol.* 13:737667. doi: 10.3389/fneur.2022.737667

**Background and Purpose:** Outcome prediction after mechanical thrombectomy (MT) in patients with acute ischemic stroke (AIS) and large vessel occlusion (LVO) is commonly performed by focusing on favorable outcome (modified Rankin Scale, mRS 0–2) after 3 months but poor outcome representing severe disability and mortality (mRS 5 and 6) might be of equal importance for clinical decision-making.

**Methods:** We retrospectively analyzed patients with AIS and LVO undergoing MT from 2009 to 2018. Prognostic variables were grouped in baseline clinical (A), MRI-derived variables including mismatch [apparent diffusion coefficient (ADC) and time-to-maximum (Tmax) lesion volume] (B), and variables reflecting speed and extent of reperfusion (C) [modified treatment in cerebral ischemia (mTICI) score and time from onset to mTICI]. Three different scenarios were analyzed: (1) baseline clinical parameters only, (2) baseline clinical and MRI-derived parameters, and (3) all baseline clinical, imaging-derived, and reperfusion-associated parameters. For each scenario, we assessed prediction for favorable and poor outcome with seven different machine learning algorithms.

**Results:** In 210 patients, prediction of favorable outcome was improved after including speed and extent of recanalization [highest area under the curve (AUC) 0.73] compared to using baseline clinical variables only (highest AUC 0.67). Prediction of poor outcome remained stable by using baseline clinical variables only (highest AUC 0.71) and did not improve further by additional variables. Prediction of favorable and poor outcomes was not improved by adding MR-mismatch variables. Most important baseline clinical variables for both outcomes were age, National Institutes of Health Stroke Scale, and premorbid mRS.

**Conclusions:** Our results suggest that a prediction of poor outcome after AIS and MT could be made based on clinical baseline variables only. Speed and extent of MT did improve prediction for a favorable outcome but is not relevant for poor outcome. An MR mismatch with small ischemic core and larger penumbral tissue showed no predictive importance.

**Keywords:** stroke, mechanical thrombectomy, outcome prediction, machine learning, MRI, perfusion imaging, mismatch

## INTRODUCTION

Mechanical thrombectomy (MT) is the most effective treatment for patients with acute ischemic stroke (AIS) due to a large vessel occlusion of the anterior circulation (1). While the average treatment effect and outcome benefit across the entire group of patients is large, outcome still differs significantly for individual patients (1). Multiple prognostic variables and their combination render individual outcome prognosis after MT difficult. For example, in a group of patients with successful and fast reperfusion, ~60% still had an unfavorable outcome (mRS 3–6 after 3 months) (2). At present, the relative importance and combination of single prognostic variables for individual outcome prediction is still a matter of debate.

One way to address this problem is to utilize artificial intelligence, in particular machine learning (ML) approaches. These models are potentially superior to conventional linear or logistic regression models as they excel at finding complex and non-linear relationships across a multitude of prognostic variables. Specially, artificial neural networks and methods of tree-boosting are promising tools in this regard (3). Recent advances have made it possible to uncover which individual prognostic variables are most important in such models (4, 5) based on a feature importance analysis.

Applying this methodology to outcome prediction after MT, multiple prognostic variables can be used representing the clinical course of stroke patients: baseline clinical variables (1), MRI variables including perfusion and infarct core (6), and finally, variables assessing the speed and extent of reperfusion (7).

The outcome and potential benefit of MT are usually assessed after 3 months with the modified Rankin Scale (mRS) in a dichotomized analysis: 0–2 is defined as favorable outcome and the remaining 3–6 as unfavorable outcome. Patients with a predicted favorable outcome will undoubtedly undergo MT. However, the remaining group of patients with unfavorable outcome is highly heterogeneous, ranging from outcomes of mRS of 3 (moderate disability) to 6 (death).

Therefore, it may also be reasonable to find prognostic factors for a poor outcome (8) (severe disability or death after 3 months, with an mRS score of 5 or 6). In those patients, withholding treatment could be discussed.

Therefore, in the presented work we used ML to predict outcome after MT directly comparing two different dichotomization paradigms: for favorable (mRS 0–2 vs. 3–6) and poor outcome (mRS 5 and 6 vs. 1–4) with multiple prognostic variables grouped in three sets: baseline clinical, MRI-derived, and MT-associated variables.

## METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

The study protocol for this retrospective analysis of our prospectively established stroke database was approved by the ethics committee of Heidelberg University and patient-informed consent was waived.

## Patients

We identified patients with AIS due to an occlusion of the middle cerebral artery in the M1 or M2 segment or the distal terminus of the internal carotid artery who were treated with MT between 03/2009 and 09/2018; 95/210 patients were treated between 2010 and 2013, and the remaining 116/210 between 2014 and 2018. Between these two groups, there was no significant outcome difference after 3 months (Mann–Whitney test,  $p = 0.83$ ).

Patients were treated at a single center (University Hospital Heidelberg). The attending neurologist and interventional neuroradiologist decided on treatment on a case-by-case basis. Only patients with a completed MRI protocol and outcome assessment at 3 months were included.

Baseline clinical and imaging parameters are given in **Table 1**. Individual patient outcome was the score on the mRS (9) at 3 months assessed by a standardized interview (unblinded investigator per phone call or a personal letter to the patient). The mTICI was used to grade recanalization on final angiographic images (10). A score of mTICI 2b or better on final angiogram was regarded as successful reperfusion.

## MRI Protocol

In a routine clinical setting, MR images were acquired on 3 Tesla MRI systems (Magnetom Verio, TIM Trio and Magnetom Prisma; Siemens Healthcare, Erlangen, Germany). Imaging protocol included diffusion-weighted, FLAIR, susceptibility

**Abbreviations:** MT, mechanical thrombectomy; AIS, acute ischemic stroke; LVO, large vessel occlusion; mRS, modified Rankin Scale; DWI, diffusion weighted imaging; MRI, magnetic resonance imaging; Tmax, time-to-maximum; mTICI, modified treatment in cerebral ischemia; NIHSS, National Institutes of Health Stroke Scale; GLM, generalized linear model; SVMC, Support Vector Machine Classifier; NB, Naive Bayes; MLP, Multilayer Perceptron; AUC, area-under-the-curve; ROC, receiver-operating-characteristic; SHAP, Shapley Additive Explanations; VIF, variance inflation factor.

**TABLE 1** | Prognostic variables (features).

		All patients (n = 210)	Favorable outcome (n = 83/210)	Poor outcome (n = 49/210)
<b>Set A: Baseline clinical variables</b>				
Time from stroke onset to MR-imaging (time in minutes)	Median/IQR	260 (126–569)	257 (123–552)	219 (109–526)
Wake up stroke	%	85 (40%)	31 (37%)	21 (43%)
Age (years)	Median/IQR	72 (59–78)	69 (53–75)	76 (68–81)
Sex	Male/female	92/118	39/44	21/28
Diabetes	%	37 (18%)	7 (8%)	14 (29%)
Hypertonia	%	131 (62%)	47 (57%)	37 (76%)
Coronary heart disease	%	36 (17%)	11 (13%)	12 (24%)
Arrhythmia/atrial fibrillation	%	77 (37%)	22 (27%)	24 (49%)
Hyperlipidemia	%	62 (30%)	23 (28%)	16 (33%)
NIHSS scale at admission (0–42)	Median/IQR	16 (12–20)	15 (10–20)	20 (15–30)
mRS pre-stroke (0–5)	Median/IQR	0 (0–1)	0 (0–1)	1 (0–2)
i.v. Thrombolysis	%	132 (63%)	58 (70%)	29 (59%)
<b>Set B: Magnetic resonance imaging derived variables</b>				
ADC lesion volume (ml)	Median/IQR	14 (8–30)	16 (7–32)	14 (9–27)
Tmax lesion volume (ml)	Median/IQR	78 (39–140)	73 (36–121)	105 (60–168)
Mismatch ratio (Tmax lesion volume/ADC lesion volume)	Median/IQR	4.6 (2.3–8.4)	4.2 (2.1–8.4)	5.3 (2.9–13.3)
Occlusion distal carotid artery	%	12 (6%)	7 (8%)	2 (4%)
Occlusion carotid terminus	%	33 (16%)	14 (17%)	7 (14%)
Occlusion M1 segment middle cerebral artery	%	131 (62%)	49 (59%)	32 (65%)
Occlusion M2 segment middle cerebral artery	%	36 (17%)	14 (17%)	7 (14%)
<b>Set C: Thrombectomy associated variables</b>				
Final mTICI score (TICI 3 and 2b)	%	154 (73%)	73 (88%)	26 (53%)
Time from stroke onset to final mTICI score	Median/IQR	492 (330–787)	526 (319–853)	434 (319–721)

Prognostic variables were grouped in three distinct sets: Baseline clinical variables (A), MRI-derived mismatch variables (B), and mechanical thrombectomy-associated variables (C). Variables are given for all patients and the two subgroups of patients with favorable outcome (mRS 0–2) and with poor outcome (mRS 5 and 6) after 3 months.

weighted and T2-weighted sequences, non-contrast time-of-flight, and contrast-enhanced angiography as well as dynamic susceptibility contrast perfusion-weighted imaging. The imaging protocol has been published previously and is included in the **Supplementary Material** (11).

## Image Post-processing

All image analysis was performed blinded to clinical outcome. Diffusion-weighted imaging (DWI) and perfusion MRI images were post-processed with Olea Sphere<sup>®</sup> (Olea Medical<sup>®</sup>, La Ciotat, France). ADC maps were automatically calculated from DWI images with different *b*-values. For perfusion imaging, automatic motion correction was applied. The arterial input function (AIF) was detected automatically. In two cases, the automatically detected AIF was manually corrected. Tmax maps were calculated using a block-circulant singular-value decomposition (cSVD) deconvolution algorithm. Diffusion lesion volumes [ADC value threshold of  $\leq 620 \times 10^{-6} \text{ mm}^2/\text{s}$  (12)] and Tmax lesion volumes [Tmax threshold  $\geq 6 \text{ s}$  (13)] were segmented semiautomatically and manually corrected for artifacts by a neuroradiologist (MM) with more than 6 years of experience in stroke imaging.

## ML Framework

For the training of the ML models, we utilized a publicly available ML framework for predictive modeling. The program code is available on Github (<https://github.com/prediction2020/explainable-predictive-models>). Details on the technical implementation have been described in open-access publications previously (5).

## Definition of Prognostic Paradigms

We defined two distinct prognostic paradigms: For the first paradigm **I**, all patients included in the study were dichotomized in favorable outcome with an mRS of 0, 1, or 2 at 3 months vs. the remaining with mRS 3–6. For the second paradigm **II**, again all patients were included and dichotomized but in poor outcome, defined as mRS 5 or 6 at 3 months vs. the remaining with mRS 0–4. The dichotomized mRS was used as a label for the ML analysis.

## Prognostic Variables for Input Feature Definition

We grouped prognostic variables in three distinct sets (**Table 1**): Baseline clinical variables (A), MRI-derived variables (B), and

thrombectomy-related variables (C). The mTICI score as a measure for the success of reperfusion was dichotomized for the final analysis (mTICI 2b/3 vs. 0–2a). The prognostic variables included in the three sets were used as input features for the ML analysis. Mismatch ratio was not included as an independent feature because it is derived from the ADC and TMax lesion volumes and would be redundant information. Target mismatch was defined according to the EXTEND-IA study (6) with an infarct core of <70 ml on ADC maps, a ratio of TMax Lesion volume to ADC lesion volume of 1.2 or higher and an absolute mismatch volume of 10 ml or more.

We defined three distinct prediction scenarios with the different sets of prognostic variables: Prediction with baseline clinical variables only (A), with baseline clinical and MRI variables combined (A+B), and finally with all baseline clinical, MRI and thrombectomy-associated variables combined (A+B+C).

For each of the three scenarios, we trained models for both prognostic dichotomization paradigms I and II with favorable (mRS  $\leq 2$ ) and poor outcome (mRS 5 or death), respectively. This yielded six different scenarios in total (see **Figure 1**).

Multicollinearity was estimated using the variance inflation factor (VIF) (5).

## Applied Algorithms

We utilized all seven available ML algorithms from the framework to provide a comprehensive coverage of various ML methods. The more traditional techniques were represented by three algorithms: A generalized linear model (GLM), which for dichotomous outcomes is equivalent to a plain logistic regression, and two regularized variants, a Lasso algorithm with L1 regularization and an ElasticNet with L1 and L2 regularization.

Further, ML algorithms included tree boosting (Catboost implementation), a Support Vector Machine Classifier (SVMC), Naive Bayes (NB), and a Multilayer Perceptron (MLP).

## Model Training and Validation

The data comprising the given clinical parameters and outcomes were randomly split into training and test sets in a corresponding 4:1 ratio. Due to slight imbalance with respect to the outcome measures (127/210 patients with favorable outcome in paradigm I and 49/210 patients with poor outcome in paradigm II), random sub-sampling of the majority class was employed for the training sets. Test sets were stratified to follow the original imbalanced ratio to represent real distribution of our patient outcomes in model testing. In total, there were only 11 missing values in the data set. Missing values were imputed by mean/mode imputation. Non-categorical features—both in training and test sets—were standardized to zero-mean and unit variance based on training set statistics. Models were trained and best parameters were selected using 10-fold cross validation over an extensive grid of hyperparameters for each model. Parameter ranges were initially taken from the public repository referenced under the heading “ML framework”, and then refined taking run times of experiments into consideration. The used ranges are included in **Supplementary Material**. The whole process was repeated

200 times (shuffles) to account for dependence on the random procedure of train and test splits.

## Performance Assessment

For performance measures, we report results as the median over the test sets of the 200 shuffles. Model performance was primarily assessed by area-under-the-curve (AUC) measure resulting from receiver-operating-characteristic (ROC) analysis. Accuracy, balanced class accuracy, precision, recall, f1 score, negative predictive value, and specificity measures for each model are included in the **Supplementary Material**. Statistical significance of the difference between model performances on the respective variable sets was determined by the Wilcoxon signed-rank test at a confidence level of 5%.

## Explainability Assessment

We used SHapley Additive exPlanations (SHAP) scores to rate the importance of included clinical features for all seven models. More detailed explanation of the technique can be found in (14). The absolute values of importance scores on test sets were scaled to unit norm to yield comparable measures for all models, and then rescaled to the range of [0, 1] so that importance scores for a certain model sum to 1. Finally, mean and standard deviation across the 200 shuffles for stability and robustness were calculated and reported as the final importance rating.

## RESULTS

### Patients

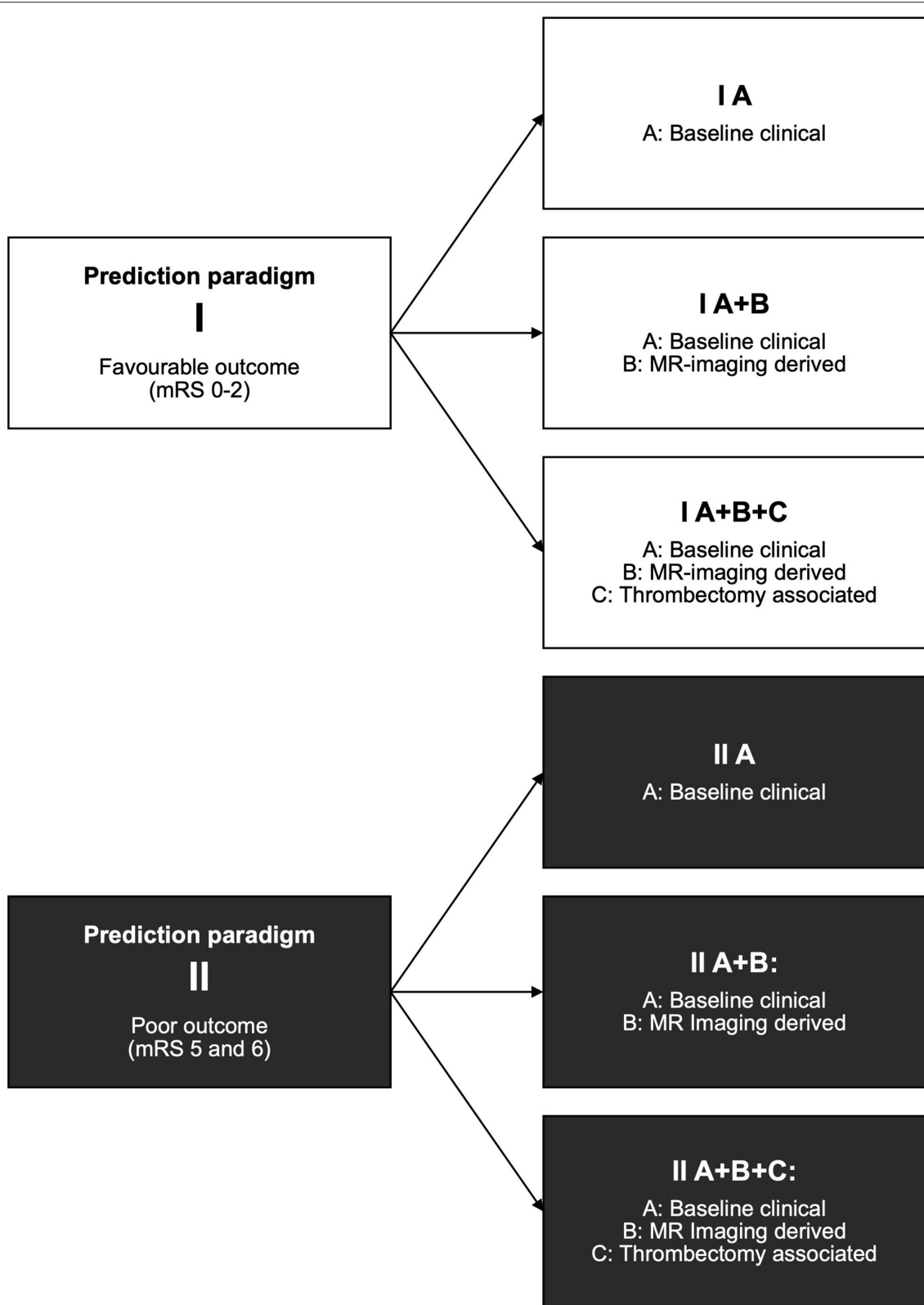
In total, 236 patients met the inclusion criteria, and 26 patients were excluded due to motion artifacts on MRI images or because no accurate AIF could be obtained, resulting in the final number of 210 patients. Median mismatch ratio was 4.6 (2.3–8.4). In 154 patients (73%), successful reperfusion (TICI 3 or 2b) could be achieved. Median time to TICI was 492 min. In prediction paradigm I, 83/210 patients (39%) had a favorable outcome (mRS 0–2). In prediction paradigm II, 49/210 patients (23%) had a poor outcome (mRS 5–6).

In this study, 168/210 patients (80%) had a target MRI mismatch [according to the EXTEND-IA study criteria (6)]; 5/210 patients (2%) had no ischemic core and 11/210 patients (5%) without target mismatch had a small ischemic core of 10 ml or less.

In the multicollinearity analysis, VIF values were below 5 for all scenarios using the predictive variable sets A and A+B. For the two scenarios I A+B+C and II A+B+C, time from stroke onset to final TICI and time from onset to MRI raised to values  $\sim 9.9$ , indicating stronger multicollinearity for these features. We did not recognize a harmful level of multicollinearity in any of the variable sets; thus, no features were eliminated.

### Prediction Models

The specific AUC results for the total of six prediction scenarios, each examined with seven algorithms are presented



**FIGURE 1 |** Prediction paradigms and resulting scenarios. For each paradigm, all patients included in the study were dichotomized: Paradigm I for favorable outcome with mRS 0–2 at 3 months (vs. the remaining 3–6) and paradigm II for poor outcome with mRS 5 and 6 (vs. the remaining 0–4). For the prediction scenarios, three sets of prediction variables A, B, and C were consecutively added. For an overview of prediction variables included in the sets, see **Table 1**. The combination of each of the three prediction variable sets and two prediction paradigms yielded six distinct scenarios.

**TABLE 2** | Models for favorable outcome (paradigm I).

Scenario	GLM	Lasso	ElasticNet	Catboost	MLP	SVMC	Naive bayes
I A	0.65	0.65	0.6	<b>0.67</b>	<b>0.67</b>	0.6	0.65
I A+B	0.62*	0.64*	0.6	<b>0.64*</b>	0.64*	0.57	0.63*
I A+B+C	0.71*+	0.71*+	0.68*+	<b>0.73*+</b>	0.7*+	0.67*+	0.69*+

The prediction variable sets (Table 1) were used to predict a favorable outcome (mRS dichotomized as 0–2 vs. 3–6). The addition of thrombectomy-associated variables (set C) leads to a noticeable improvement of all machine learning models. Highest AUC results for each variable set are marked in bold. Confidence intervals for all models are included in the **Supplementary Material**. Statistically significant difference in model performance between variable sets are marked with \* and + to signal difference from A and from A+B, respectively. Significance was determined by a value of *p* lower than 0.05, resulting from the Wilcoxon signed-rank test.

**TABLE 3** | Models for poor outcome (paradigm II).

Scenario	GLM	Lasso	ElasticNet	Catboost	MLP	SVMC	Naive bayes
II A	0.67	0.7	0.64	0.7	<b>0.71</b>	0.59	0.69
II A+B	0.65*	<b>0.7</b>	0.62*	<b>0.7</b>	0.69	0.57	0.65*
II A+B+C	0.68+	0.71	0.65+	<b>0.73*+</b>	0.7	0.65*+	0.66*

The prediction variable sets (Table 1) were used to predict poor outcome (mRS dichotomized as 5 and 6 vs. 0–4). In contrast to the favorable outcome paradigm, the addition of thrombectomy-associated variables (set C) did not lead to relevant improvements in the performance of machine learning models. Only the Catboost model profited slightly in a clinically relevant AUC range. Highest AUC results for each variable set are marked in bold. Confidence intervals for all models are included in the **Supplementary Material**. Statistically significant difference in model performance between variable sets are marked with \* and + to signal difference from A and from A+B, respectively. Significance was determined by a value of *p* lower than 0.05, resulting from the Wilcoxon signed-rank test.

in Table 2 for paradigm I with favorable outcome and in Table 3 for paradigm II with poor outcome, respectively. The results for the additional performance measures are given in **Supplementary Material**.

For the first scenario with baseline clinical variables only (I A and II A), prediction was slightly better for poor outcome (II A) than for favorable outcome (I A). The smallest difference in AUC between I A and II A was 0.02 for GLM and the largest 0.05 for Lasso logistic regression. Only SVMC showed comparable results.

Adding MRI-derived parameters (scenario I A+B and II A+B) did not change the prediction performance for both paradigms. This was consistent for all algorithms in both scenarios I A+B and II A+B.

Finally, adding thrombectomy-associated parameters—extent and speed of recanalization—I A+B+C and II A+B+C improved the prediction performance noticeably across many algorithms for the favorable outcome paradigm (I A+B+C). Prediction for the poor outcome paradigm with all variables (II A+B+C) remained approximately stable; only the Catboost and SVMC algorithm showed a slight improvement (AUC increase of 0.03, 0.08, respectively, compared to II A).

To summarize, prediction for the poor outcome paradigm II remained comparable on a relatively high level across all three prediction scenarios (II A, II A+B, II A+B+C). Contrariwise, prediction for the favorable outcome paradigm I improved noticeably when thrombectomy-associated parameters were added (I A+B+C). The final performance for the last scenario with all predictive variables included (I A+B+C and II A+B+C) was comparable for both the favorable and poor paradigms.

## Feature Importance Ranking

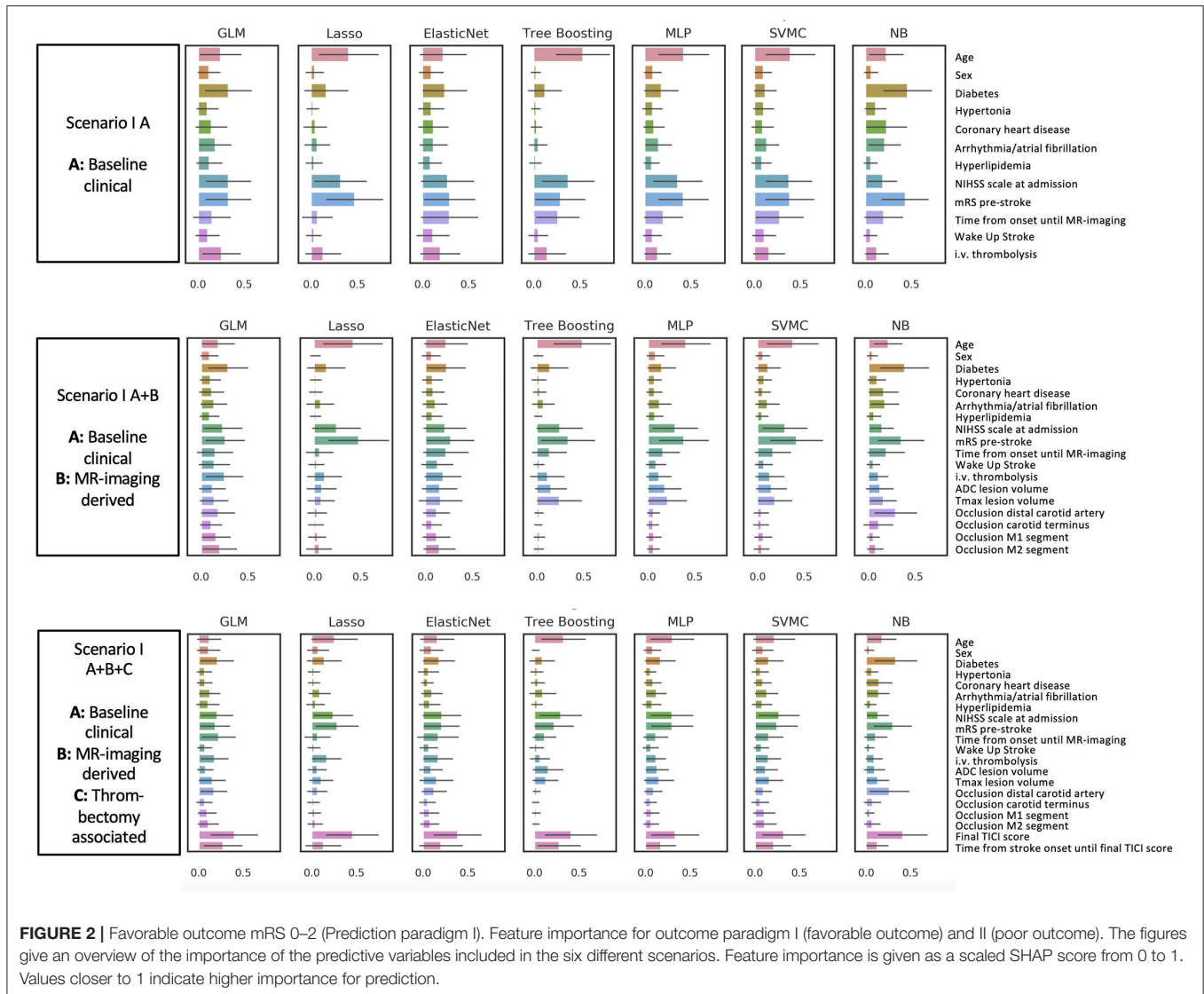
Feature importance values for each scenario and each algorithm are displayed in **Figures 2, 3**.

For the favorable outcome paradigm I (Figure 2), the most relevant variables across all algorithms for the first scenario with the baseline clinical variables only (I A) were age, pre-stroke mRS, National Institutes of Health Stroke Scale (NIHSS) at admission, and time from stroke onset to MRI. Adding MRI-derived mismatch parameters and the site of occlusion (I A+B), the Tmax volume for hypoperfused tissue was of higher relevance, while the ADC volume was of moderate importance. However, this was only visible in some models, among them the tree boosting model (the best performing model in the favorable outcome paradigm I). For models with all variables combined (I A+B+C), the mTICI score became the most dominant parameter in all models. Also, time from stroke onset to final TICI score was assigned high importance by the majority of models.

For the poor outcome prediction paradigm (Figure 3), we found a similar pattern. Age, premorbid mRS, and the baseline NIHSS were the most relevant features in the model with baseline variables only (II A). However, i.v. thrombolysis and risk factors such as diabetes played a smaller role compared to the favorable outcome paradigm (I A). For the models with additional MR-mismatch parameters (II A+B), both ADC and Tmax volume were of less importance than in the favorable outcome paradigm (I A+B). In the third scenario with all variables included (II A+B+C), the baseline variables from the first scenario (II A) remained important. Additionally, the mTICI score was relevant in most models, however not as relevant as compared to the favorable outcome prediction (I A+B+C).

To summarize, for the prediction of either good or poor outcome, age, premorbid mRS, and baseline NIHSS were





important, with mTICI score as an additional relevant feature from the third scenario. ADC and Tmax volume were more important for the favorable than for the poor outcome paradigm. Information about i.v. thrombolysis was only important for the favorable outcome paradigm.

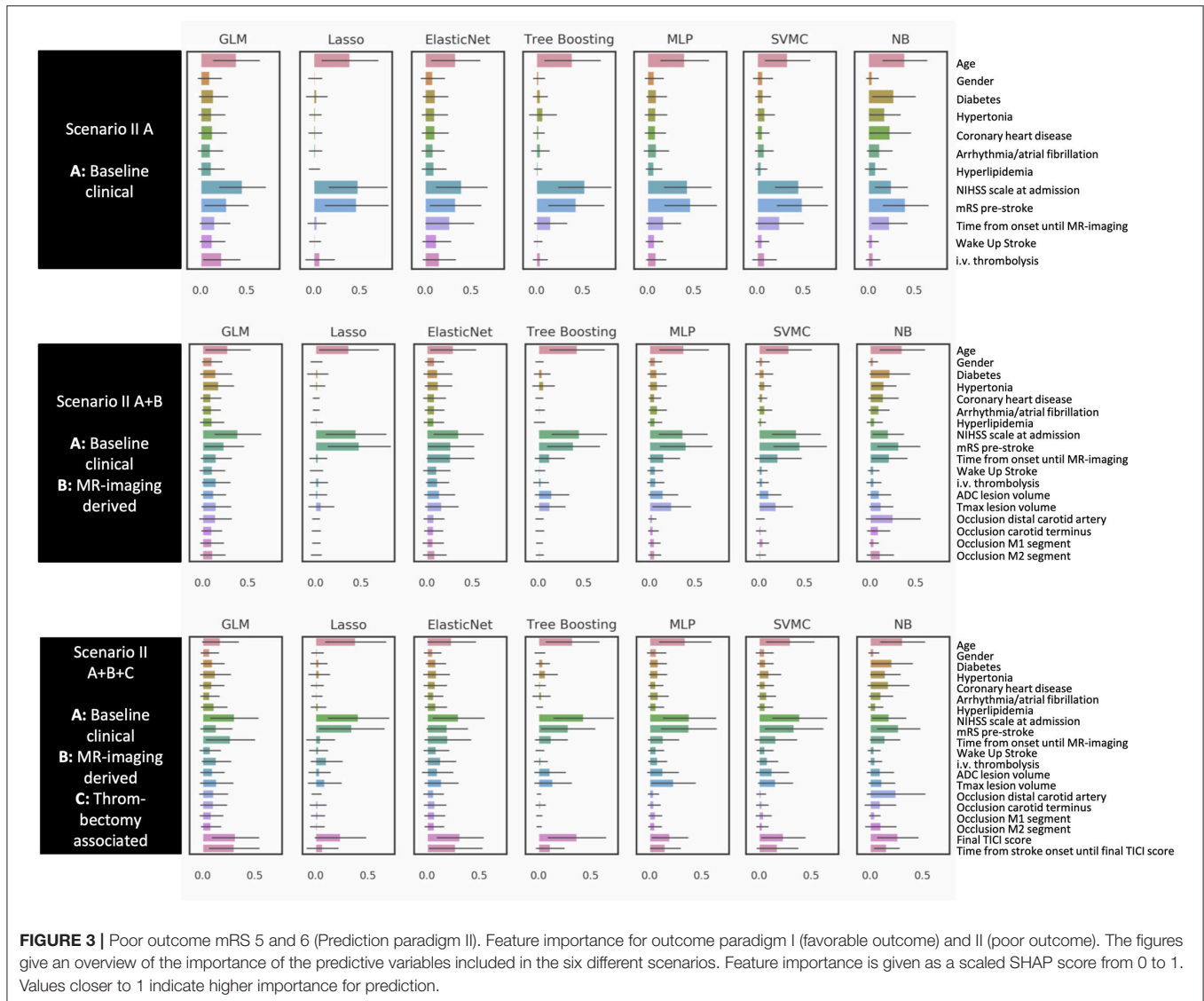
## DISCUSSION

In this study, we examined ML-based outcome prediction models for patients with stroke who underwent MT. We compared prediction of poor outcome (mRS 5 or 6 vs. 0–4) and favorable outcome (mRS 0–2 vs. 3–6) measured at 3 months post stroke. These prediction paradigms have direct implications for clinical decision-making by predicting an outcome of no or only slight disability on the one hand and severe disability or death on the other. In particular, the definition of favorable outcome is generally accepted and was applied in large prospective studies.

We chose different combinations of prognostic variables that were deliberately limited to those most commonly used in stroke practice and literature (1, 6, 7) and most accessible in clinical decision-making, especially under time constraints as encountered in clinical practice. We found considerable differences between the two outcome paradigms.

Our main finding suggests that prediction of poor outcome may possibly be based on clinical baseline variables only and set a rather high benchmark in the first prediction scenario. The predictive performance did not improve by adding target MR-mismatch and recanalization-related parameters. In contrast, prediction of favorable outcome did improve significantly by adding speed and extent of recanalization compared to using baseline clinical variables only.

In contrast to previous studies, the main strength of the presented work is the direct comparison of two different outcome prediction paradigms. The choice of how to dichotomize mRS for outcome prediction has clinical relevance: The standard



dichotomization with 0–2 vs. 3–6 as used in the large randomized clinical trials puts an emphasis on patients with favorable outcomes. Dichotomizing mRS 0–4 vs. 5–6 focuses on patients with a very high degree of dependency or mortality after MT.

Our exploratory analysis revealed that patients with expected poor outcome could already be captured by clinical baseline variables before thrombectomy. The prediction scenario with baseline clinical variables only was already close to the final performance when mismatch and recanalization information was added. In those patients, withholding MT based on the clinical baseline variables could be the consequence; however, these implications should be verified in larger, prospective studies (8) and within new clinical data sets.

This was in contrast to the prediction of favorable outcome, where the prediction with baseline clinical variables only

was lower but could be considerably improved by adding information about the speed and extent of reperfusion. However, because of the retrospective and thus exploratory nature of our study, our results should be interpreted with caution with regards to clinical treatment decisions. Nonetheless, they warrant further exploration in prospective studies to confirm our findings. Within such prospective data, the proposed paradigms could then be used to estimate individualized chances for either poor or favorable outcome before and after therapy. A similar approach using two models for an individualized prognosis of the same outcome variable is proposed by Debs et al. (15).

This performance pattern was accompanied by complimentary information from the feature importance analysis: For poor outcome, age, stroke severity (NIHSS) and degree of disability before stroke (premorbid mRS before stroke), and the time from stroke onset to imaging were the most

important baseline variables. For favorable outcome, the most important predictive variables were also age, NIHSS at onset, and the mRS before stroke, but additionally, the speed and extent of recanalization (mTICI score and the time from stroke onset to the final mTICI score) were paramount. NIHSS was a more important predictor for final outcome than the ischemic lesion volume before therapy. This could be due to lesion location in eloquent brain regions where a smaller infarct causes comparably more severe clinical symptoms. Also, the final ischemic lesion volume after therapy might improve prediction but was not available to be included in our models.

Across all models, the maximum predictive value was an AUC of 0.73 for both favorable and poor outcome with regards to model performance and feature rankings. Our findings are comparable to previous studies applying ML algorithms: Hammam et al. (16) found a similar prediction for favorable outcome for patients with MT which did not considerably improve by adding MR-mismatch and other imaging-derived parameters. Other studies applied ML for outcome prediction with baseline CT imaging: Brugnara et al. (17) found an AUC of 0.85 for the prediction of favorable outcome only after adding information about infarct size after thrombectomy. Ramos et al. (18) did evaluate prediction for poor outcome with a multitude of clinical baseline parameters and CT-derived imaging features in a much larger cohort. Not including mismatch variables, their highest AUC was 0.81. Van Os et al. similarly showed a considerable improvement for prediction by adding treatment-associated variables in a study including CT imaging (19).

Interestingly, prediction for either poor or favorable outcome did not improve by including MR-mismatch variables. These results need to be interpreted together with the characteristics of the cohort: Most patients included had a target mismatch as defined in the inclusion criteria for the EXTEND-IA study (6) with an infarct core of <70 ml on ADC maps and a comparably larger volume of hypoperfused tissue on Tmax maps with a mismatch volume ratio of 1.2 or higher. While the treatment effect of MT is maintained even in patients with larger infarct cores (20), individual patients with a target mismatch still have poor outcome. Therefore, it is intriguing that poor outcome prediction in our study was possible based on clinical baseline variables only: For those patients, the potential predictive value of MR mismatch variables could be already encoded in the clinical baseline information. A similar conclusion can be drawn for patients with favorable outcome: improved prediction was much more dependent on speed and the extent of recanalization than on MR mismatch. However, this does not preclude the possibility that a target mismatch is still a valid selection criterion for patients undergoing MT. Our sample does not allow a conclusion about the potential predictive value of patients without MR mismatch. Prediction models could be improved by including patients who underwent MT without a target mismatch profile and larger infarct cores (21).

Despite these findings, the overall performance of the ML models tested in our study could be improved. Considering the potential power of ML algorithms to extract patterns, our findings suggest that important variables for outcome prediction might not be included in today's clinical decision-making. It

is conceivable that there are so far unknown or undetectable variables. This warrants further studies including more and new prediction variables and biomarkers as well as direct integration of multimodal imaging and clinical information.

A deep learning model including raw imaging data and not derived variables might extract further, previously unknown predictive information. For example, these models might be able to account for inherent errors in the definition of infarct core (22) or individual susceptibility of brain tissue (21).

Finally, our results show that the predictive value can differ significantly between two different dichotomization paradigms or different "cutoffs" (mRS 0–2 vs. 3–6 and mRS 5/6 vs. 0–4). Unnecessary dichotomization of the mRS can be suboptimal (23). Researchers and clinicians should be aware that there are relevant differences between dichotomization paradigms. Defining more accurate outcome or premorbidity scores might improve future prediction models.

Our study has some limitations. It is based on a relatively small and retrospective patient cohort. Thus, our results must be understood as an exploratory analysis for future research.

Our data reach back to 2009. This might have influenced outcome due to improvements in thrombectomy technique and accelerated workflows. However, there was no significant outcome difference between patients treated 2009–2013 vs. 2014–2018.

## CONCLUSION

Our results suggest that a prediction of poor outcome (mRS of 5 or 6) after MT can be based on clinical baseline variables only. Speed and extent of thrombectomy did not seem to influence poor outcome but were important for favorable outcome (mRS 0–2). The predictive value of a target MR mismatch with smaller infarct core and larger penumbra was not relevant and could be already captured by clinical baseline variables. However, our sample does not allow a conclusion about the predictive value in patients without target MR mismatch.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Heidelberg University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

MMu, AP, CW, MMö, PR, SN, and MB acquired the data and organized the database. VM, AH, EZ, and DF performed and developed the machine learning and statistical

analysis. MMu wrote the first draft of the manuscript. VM, AH, and DF wrote sections of the manuscript. All authors contributed to the conception, design of the study, manuscript revision, read, and approved the submitted version.

## REFERENCES

- Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
- Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA*. (2016) 316:1279–89. doi: 10.1001/jama.2016.13647
- Livne M, Boldsen JK, Mikkelsen IK, Fiebach JB, Sobesky J, Mouridsen K. Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke. *Stroke*. (2018) 49:912–8. doi: 10.1161/STROKEAHA.117.019440
- Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. (2020) 20:310. doi: 10.1186/s12911-020-01332-6
- Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. *PLoS ONE*. (2020) 15:e0231166. doi: 10.1371/journal.pone.0231166
- Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. (2015) 372:1009–18. doi: 10.1056/NEJMoa1414792
- Jayaraman MV, Grossberg JA, Meisel KM, Shaikhouni A, Silver B. The clinical and radiographic importance of distinguishing partial from near-complete reperfusion following intra-arterial stroke therapy. *Am J Neuroradiol*. (2013) 34:135–9. doi: 10.3174/ajnr.A3278
- Goyal M, Almekhlafi MA, Cognard C, McTaggart R, Blackham K, Biondi A, et al. Which patients with acute stroke due to proximal occlusion should not be treated with endovascular thrombectomy? *Neuroradiology*. (2019) 61:3–8. doi: 10.1007/s00234-018-2117-y
- Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. *Stroke*. (1999) 30:1538–41. doi: 10.1161/01.STR.30.8.1538
- Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, Von Kummer R, Saver JL, et al. Recommendations on angiographic revascularization grading standards for acute ischemic stroke. *Stroke*. (2013) 44:2650–63. doi: 10.1161/STROKEAHA.113.01972
- Potreck A, Loebel S, Pfaff J, Østergaard L, Mouridsen K, Radbruch A, et al. Increased volumes of mildly elevated capillary transit time heterogeneity positively predict favorable outcome and negatively predict intracranial hemorrhage in acute ischemic stroke with large vessel occlusion. *Eur Radiol*. (2019) 29:3523–32. doi: 10.1007/s00330-019-06064-4
- Purushotham A, Campbell BCV, Straka M, Mlynash M, Olivot JM, Bammer R, et al. Apparent diffusion coefficient threshold for delineation of ischemic core. *Int J Stroke*. (2015) 10:348–53. doi: 10.1111/ijvs.12068
- Zaro-Weber O, Moeller-Hartmann W, Siegmund D, Kandziora A, Schuster A, Heiss WD, et al. MRI-based mismatch detection in acute ischemic stroke: optimal PWI maps and thresholds validated with PET. *J Cereb Blood Flow Metab*. (2017) 37:3176–83. doi: 10.1177/0271678X16685574
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17*. Curran Associates Inc. (2017). p. 4768–77.
- Debs N, Cho TH, Rousseau D, Berthezène Y, Buisson M, Eker O, et al. Impact of the reperfusion status for predicting the final stroke infarct using deep learning. *Neuroimage Clin*. (2021) 29:102548. doi: 10.1016/j.nicl.2020.102548
- Hamann J, Herzog L, Wehrli C, Dobrocky T, Bink A, Piccirelli M, et al. Machine learning based outcome prediction in stroke patients with MCA-M1 occlusions and early thrombectomy. *Eur J Neurol*. (2021) 28:1234–243. doi: 10.1111/ene.14651
- Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke*. (2020) 51:3541–51. doi: 10.1161/STROKEAHA.120.030287
- Ramos LA, Kappelhof M, van Os HJA, Chalos V, Van Kranendonk K, Kruyt ND, 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
- van Os HJA, Ramos LA, Hilbert A, Van Leeuwen M, Van Walderveen MA, Kruyt ND, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: potential value of machine learning algorithms. *Front Neurol*. (2018) 9:784. doi: 10.3389/fneur.2018.00784
- Campbell BCV, Majoie CBLM, Albers GW, Menon BK, Yassi N, Sharma G, et al. Penumbra imaging and functional outcome in patients with anterior circulation ischaemic stroke treated with endovascular thrombectomy versus medical therapy: a meta-analysis of individual patient-level data. *Lancet Neurol*. (2019) 18:46–55. doi: 10.1016/S1474-4422(18)30314-4
- Goyal M, Menon BK, Almekhlafi MA, Demchuk A, Hill MD. The need for better data on patients with acute stroke who are not treated because of unfavorable imaging. *AJNR Am J Neuroradiol*. (2017) 38:424–5. doi: 10.3174/ajnr.A5094
- Goyal M, Ospel JM, Menon B, Almekhlafi M, Jayaraman M, Fiehler J, et al. Challenging the ischemic core concept in acute ischemic stroke imaging. *Stroke*. 51:3147–55. doi: 10.1161/STROKEAHA.120.030620
- Ganesh A, Luengo-Fernandez R, Wharton RM, Rothwell PM. Ordinal vs dichotomous analyses of modified rankin scale, 5-year outcome, and cost of stroke. *Neurology*. (2018) 91:e1951–60. doi: 10.1212/WNL.00000000000006554

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.737667/full#supplementary-material>

**Conflict of Interest:** VM reported receiving personal fees from ai4medicine outside the submitted work. AH reported receiving personal fees from ai4medicine outside the submitted work. DF reported receiving grants from the European Commission Horizon2020 PRECISE4Q No. 777107, reported receiving personal fees from and holding an equity interest in ai4medicine outside the submitted work. There is no connection, commercial exploitation, transfer, or association between the projects of ai4medicine and the results presented in this work. SN received unrelated fees for consultancy from Brainomix and Boehringer Ingelheim, payment for lectures including service on speakers' bureaus from Pfizer, Medtronic, and Bayer AG. MB received unrelated grants from Siemens, grants and personal fees from Novartis, grants from Stryker, grants from DFG, personal fees from Merck, personal fees from Bayer, personal fees from Teva, grants and personal fees from Guerbet, personal fees from Boehringer, personal fees from Vascular Dynamics, personal fees from Grifols, and grants from the European Union, all outside the submitted work. MMö received unrelated Board Membership from Codman; consultancy from Medtronic, MicroVention,

and Stryker; payment for lectures including service on speakers bureaus' from Medtronic, MicroVention, and Stryker. PR received unrelated grants for consultancy from Boehringer and lecture fees from Bayer, Boehringer Ingelheim, BMS, Daichii Sankyo, and Pfizer.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Mutke, Madai, Hilbert, Zihni, Potreck, Weyland, Möhlenbruch, Heiland, Ringleb, Nagel, Bendszus and Frey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*