

## TITLE PAGE

Title

iSPAN: Improved prediction of outcomes post thrombectomy with Machine Learning

KEYWORDS

Machine Learning; Outcomes; Radiology, Interventional; Stroke; Thrombectomy

## AUTHORS

**Brendan S Kelly MB 1-6 (Corresponding Author)**, Prateek Mathur<sup>2</sup> MSc, John Duignan MB, Sarah Power MB, Edward H Lee PhD 6, Yuhao Huang MD 6, Silvia D Vaca MD 6, Laura M Prolo MD 6, Kristen W Yeom MD 6, Aonghus Lawlor PhD 2, Ronan P Killeen MB 1

## Affiliations

**1. St Vincent's University Hospital, Dublin, Ireland**

Brendan Kelly, Simon M. Clifford, Gerard M. Healy & Ronan P. Killeen

**2. Insight Centre for Data Analytics, UCD, Dublin, Ireland**

Brendan Kelly & Aonghus Lawlor

**3. Wellcome Trust – HRB, Irish Clinical Academic Training, Dublin, Ireland**

Brendan Kelly, Conor Judge & Stephanie M. Bollard

**4. School of Medicine, University College Dublin, Dublin, Ireland**

Brendan Kelly & Stephanie M. Bollard

**5. Beaumont Hospital, Dublin, Ireland**

Brendan Kelly, John Duignan, John Thornton, Sarah Power

**6. Lucille Packard Children's Hospital at Stanford, Stanford, CA, USA**

Kristen W. Yeom

**Corresponding Author: Brendan S Kelly**

**BSK ORCID: 0000-0002-3449-8017**

e-mail address [brendanskelly@me.com](mailto:brendanskelly@me.com) telephone 00353 1 221 4000

fax (00353 1) 221 3394

## Funding

This work was performed within the Irish Clinical Academic Training (ICAT) Programme, supported by the Wellcome Trust and the Health Research Board (Grant No. 203930/B/16/Z), the Health Service Executive National Doctors Training and Planning and the Health and Social Care, Research and Development Division, Northern Ireland and the Faculty of Radiologists, Royal College of Surgeons in Ireland. This research was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289\_P2.

## Key words:

Machine Learning; Outcomes; Radiology, Interventional; Stroke; Thrombectomy

## Key Points:

- An XGB model performed better than existing scores and other tested models for prognostication post EVT.
- It identified mTICI and number of passes as important and modifiable factors.
- Integrating these into the SPAN score (iSPAN) was not inferior to the XGB model and is generalisable and easier to use and interpret.

Figures: 2

Tables: 3

~~NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.~~

Total Word count: 4573

## Abstract

### Background:

This study aimed to develop and evaluate a machine learning model and a novel clinical score for predicting outcomes in stroke patients undergoing endovascular thrombectomy.

### Methods:

This retrospective study included all patients aged over 18 years with an anterior circulation stroke treated at a thrombectomy centre from 2010 to 2020. External validation data was obtained. The primary outcome variable was day 90 mRS  $\geq 3$ . Existing clinical scores (SPAN and PRE) and Machine Learning (ML) models were compared. A novel clinical score (iSPAN) was derived by adding an optimised weighting of the most important ML features to the SPAN and compared results.

### Results:

812 patients were initially included (397 female, average age 73), 62 for external validation. The best performing clinical score and ML model were SPAN and XGBoost (sensitivity specificity and accuracy 0.967, 0.290, 0.628 and 0.783, 0.693, 0.738 respectively). A significant difference was found overall and XGBoost was more accurate than SPAN ( $p < 0.0018$ ). The most important features were Age, mTICI and total number of passes. The addition of 11 points for mTICI of  $\leq 2B$  and 3 points for  $\geq 3$  passes to the SPAN achieved the best accuracy and was used to create the iSPAN. iSPAN was not significantly less accurate than XGBoost ( $p > 0.5$ ). In the external validation set, iSPAN and SPAN achieved sensitivity, specificity, and accuracy of (0.735, 0.862, 0.79) and (0.471, 0.897, 0.67), respectively.

Conclusions: iSPAN incorporates machine-derived features to achieve better predictions compared to existing scores. It is not inferior to the XGB model and is externally generalisable.

### Non standard abbreviations

Artificial Intelligence (AI)

Endovascular thrombectomy (EVT)

eXtreme Gradient Boosting (XGB)

K-Nearest Neighbour (KNN)

Support Vector Machine (SVM)

Machine Learning (ML)

Multilayer Perceptron (MLP)

Stroke Prediction using Age and NIHSS (SPAN)

The added "I" in iSPAN refers to Improved, artificial Intelligence, Insight (our research centre), and the ICAT programme (our funders)

Pittsburgh Response to Endovascular therapy (PRE)

## Introduction

Acute ischaemic stroke is a time-dependent disease that continues to be a significant cause of morbidity and mortality worldwide<sup>1-3</sup>. The advent of endovascular thrombectomy (EVT) has revolutionized acute stroke care, leading to improved patient outcomes<sup>4</sup>. As a result, there has been a marked increase in the number of patients eligible for and undergoing EVT<sup>5</sup>. However, a substantial proportion of patients do not achieve functional independence despite undergoing EVT. This makes the identification of patients who will have better outcomes post-EVT an important clinical challenge, both in terms of prospective patient selection and identification of patients at risk for poor outcome as soon as possible post thrombectomy.

To reach the agreed standard that at least 30% of patients should be functionally independent at 90 days post thrombectomy, as indicated by a 90-day modified Rankin score (mRS) of less than 3<sup>6</sup>, it is crucial to identify modifiable factors that may improve outcomes. Prognostication plays a vital role in guiding the selection of appropriate medical measures or surgical therapies for patients. Secondary data analysis of stroke databases has demonstrated the potential to identify practice-changing insights that could help improve patient selection for thrombectomy<sup>7</sup>. Several potential prognostic factors have been suggested<sup>8,9</sup>. While predictive scores have been proposed and trialled for this purpose<sup>10</sup>, existing patient selection models are often based on data from patients who have already undergone EVT, resulting in selection bias and overly optimistic predictions of good outcomes. The integration of post-EVT information into selection scores may improve their prognostic utility, driving interest in the potential application of machine learning for predicting outcomes in these patients<sup>11</sup>. However, preliminary studies in this area are limited by small sample sizes, with a recent systematic review of the topic including only 802 patients<sup>12</sup>. Further research with larger cohorts is necessary to ascertain the effectiveness of machine learning models in predicting outcomes for patients with stroke who undergo EVT.

Management of acute stroke has been disrupted by the advent of thrombolysis<sup>13</sup>, thrombectomy, and more recently by artificial intelligence in diagnostic<sup>14</sup> and interventional radiology<sup>15</sup>. Despite the potential benefits of Artificial Intelligence (AI) in clinical settings, there are several barriers to its more widespread adoption<sup>16</sup>. One notable challenge is the "black box" nature of many AI models, which can make it difficult for clinicians to understand and trust the decision-making process of these algorithms<sup>17</sup>. This lack of transparency can hinder acceptance and limit the ability to identify potential sources of error<sup>18</sup>. Additionally, integrating AI models into existing clinical software and workflows can be complex, requiring significant resources and technical expertise<sup>19</sup>. Generalisability is also a concern, as AI models trained on specific patient populations or healthcare systems may not perform as well when applied to different contexts<sup>20</sup>. Furthermore it has been shown that results of more complex stroke prediction models may not generalise to minority populations, and there have been recent calls for expansion of the pool of risk factors and improvements in modelling to mitigate against racial bias<sup>21</sup>.

This study aimed to use machine learning to identify the variables that are predictive of an unfavourable outcome (90 day Modified Rankin Score (mRS)  $\geq 3$ ) in patients with an anterior circulation ischaemic stroke selected for EVT, and to determine if these features be

combined with existing patient selection models to better predict outcome. Our hypotheses were that a machine learning model would identify patients post thrombectomy at higher risk of poor outcome and that a simpler model would be more generalisable.

## Methods

This retrospective cohort study received institutional review board approval and the need for prospective patient informed consent was waived under quality improvement grounds. This manuscript was prepared using the STROBE checklist<sup>22</sup>. All patients aged over 18 years with an anterior circulation stroke treated by our National Thrombectomy Service from 2010 to 2020 were included. All patients with an anterior circulation stroke treated at an international external tertiary referral centre between 2010 and 2018 were included for external validation. Both Pittsburgh Response to Endovascular therapy score and SPAN-100 score have been shown to be effective in similar studies and were used as performance comparison in this case<sup>23,24</sup>.

### Statistical analysis and sample size calculation

The target variable was a binary prediction of mRS at 90 days extracted from the patients clinical chart (internal) or EPIC (external) and in all cases was calculated by a staff neurologist or stroke physician independently to the study. Sample size calculation was undertaken to develop a binary prediction model<sup>25</sup>. A small optimism in predictor effect and estimates (as defined by a global shrinkage factor of  $\geq 0.9$ ) was used, as was with a small absolute difference of  $\leq 0.05$  in the model's apparent and adjusted Nagelkerke's R<sup>2</sup>. A precise estimation of the overall risk in the population was available. This showed that for a model with 4 predictive parameters a sample of at least 322 patients for development was required. Cochran's Q test was used to compare between multiple models and McNemar's test was used in cases where two models were compared. Data were divided into an 80% 20% train test split. An external validation set was obtained to assess generalisability.

### Model training, evaluation and development of new predictive score

Several machine learning models (eXtreme Gradient Boosting (XGB), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Multilayer Perceptron (MLP)) were trained and tested and compared with two existing clinical scores (Stroke Prediction using Age and NIHSS (SPAN) and Pittsburgh Response to Endovascular therapy (PRE)). 10-fold cross validation was used to assess performance of the models in the training set. Hyperparameter optimisation was employed using a grid search. The best performing optimized model was tested on a holdout set. For evaluation we computed and considered metrics including accuracy, sensitivity and specificity. Accuracy was used as the primary evaluation metric as our binary target class were relatively well balanced. Feature importance (gain based) were extracted using the `plot_importance` method in Scikit learn.

Feature importance was extracted from the best performing ML model and combined with best performing clinical score to create a novel *intelligence* clinical score. We then tested to determine non inferiority of the novel clinical score to the ML model. Finally the novel score was externally validated at an international institution and preliminary subgroup analysis was undertaken to assess for demographic effects.

## Results

A total of 812 patients were included in the initial analysis, with 397 being female and an average age of 73 (range 18-93) with 122 used for testing (62 with Day90-mRS of  $\geq 3$ ). For external validation, 63 patients were included, comprising 42 females and an average age of 74 (range 28-92) 29 with Day90-mRS of  $\geq 3$ . Complete patient demographics are presented in Table 1. The best-performing clinical score and machine learning (ML) model were SPAN (sensitivity, specificity, accuracy) (0.967, 0.290, 0.628) and XGB (0.783, 0.693, 0.738). A significant difference in accuracy was observed overall, with XGB demonstrating significantly higher accuracy than the next best predictor, SPAN ( $p < 0.0018$ ). Comprehensive results can be found in Table 2.

Feature importance of the XGB model is illustrated in Figure 1. SPAN can be calculated as (Age + NIHSS), with scores over 100 associated with poor outcomes. The most important features identified were Age, mTICI, and Total number of Passes. Given that mTICI and number of Passes are not included in SPAN, simulations were run to optimize the weight of their inclusion in the integrated SPAN (iSPAN) model (Figure 2). In the training set, adding 11 points for an mTICI of 2B or lower and a number of passes equal to or greater than 3 was found to be the optimal weighting for predicting poor outcomes. No significant difference in performance was observed between iSPAN and XGB on the internal test set ( $p > 0.5$ ) (Table 2).

In the external validation set, iSPAN and SPAN achieved sensitivity, specificity, and accuracy of (0.735, 0.862, 0.79) and (0.471, 0.897, 0.67), respectively. A comparison of the performance of SPAN and iSPAN on internal and external data is provided in Table 3. To help mitigate the propagation of racial or ethnic bias, the performance of iSPAN and SPAN on the external validation set for Caucasian patients compared to non-Caucasian patients is also presented.

## Discussion

This study used an XGB model to identify features which can predict a poor outcome post thrombectomy. We also showed that integrating the features with the highest importance with an existing clinical score had similar performance to a more complicated ML model. The results demonstrate the potential utility of machine learning algorithms, such as XGB, and the integration of machine-derived additional variables in traditional clinical scores, in improving the accuracy and generalizability of outcome predictions for patients with stroke who undergo EVT. Our results are consistent with and build upon previous literature, which has shown the potential of machine learning algorithms to enhance prognostication in the context of acute ischemic stroke and EVT<sup>12</sup>.

Prediction of outcome post thrombectomy with machine learning is a common research question in the literature<sup>11,12,26</sup>. Models such as XGB show state of the art performance on tabular data<sup>27</sup> and have been successfully applied to smaller stroke databases<sup>28</sup>. There is also hope that machine learning can add a level of personalisation to decision making around thrombectomy<sup>29</sup>. However there is a large degree in heterogeneity in these studies<sup>12,29</sup>. Many are also hampered by low sample size<sup>26,28,30-32</sup>. Indeed a systematic review of the

topic had a combined cohort of only 802 patients<sup>12</sup>, meaning that our internal cohort of 812 is a substantial contribution to the literature. Another advantage of our model is that we validated our results on an external population, unlike many other models that use internal or indeed no validation of their model<sup>26,28,30-32</sup>. iSPAN's performance metrics compare very favourably with published models especially when you consider its performance on external data.

The simplicity of iSPAN is another area where our model is different from the current literature. Even recent developments that focus strictly on interpretability suggest the use of additional algorithms to aid in explanations<sup>33</sup>. We propose the addition of mTICI and the number of passes, which are both factors well known to influence outcome<sup>7,8,30,34</sup> and an intuitive reason for their effect on outcomes. That age, mTICI and number of passes were the features identified as most important by machine learning also gives an intuition that our model is behaving as expected. Furthermore imaging factors such as ASPECTS and collaterals were weighted less highly which is concordant with the literature, however this is likely because of the inclusion criteria for thrombectomy.

The simplicity and interpretability of iSPAN offers key advantages over more complex machine learning models. Although a more complicated model might exhibit higher overall accuracy, a "black box" nature can impede clinicians' and patients' understanding and trust in the decision-making process<sup>35</sup>. In contrast, iSPAN integrates additional variables into the traditional SPAN score, making it more easily calculable, comprehensible and clinically applicable<sup>36</sup>. iSPAN's ease of use and interpretability potentially renders it more appealing, as it allows for better insight into factors affecting a patient's predicted outcome<sup>35</sup>. By providing a transparent scoring system, iSPAN promotes trust and confidence among healthcare professionals and patients and families, fostering its adoption in routine clinical practice<sup>17,35,37</sup>. Moreover, iSPAN's simplicity (the addition of four numbers and comparing to 100) enables seamless integration into existing workflows, mitigating barriers often associated with complex AI models and legacy software systems<sup>36</sup>, crucial when "time is brain". This effectively instant feedback is another advantage over other proposed models that require additional information over days or weeks<sup>28</sup>.

There are two main avenues of clinical utility. First iSPAN can be used to give immediate prognostic information post thrombectomy to identify those patients at higher risk for a poor outcome. It is possible that such patients could benefit from more intense medical management. The second is the conformation of the importance of both maximising the mTICI while minimising the number of passes. Balancing these two factors is an important factor for interventional neuroradiologists to consider. It is also possible that this information prospectively could aid in selection of patients, for example where anatomical factors might make a higher number of passes more likely, but this would need prospective evaluation.

#### Limitations and bias

There is a strong selection bias in our data as it includes only to patients who were selected for thrombectomy. As such it should only be applied to similar cohorts until further research shows prospective utility. No racial or ethnic data is available for the internal cohort but it is known that the overall population undergoing thrombectomy is overwhelmingly "White

Irish” due to local population demographics. Performance on Caucasian compared to non-Caucasian patients is reported where available (on the external validation set) to help to mitigate against any further propagation of racial or ethnic bias. Many existing clinical stroke scores are known to perform worse on minority populations<sup>21</sup>. While no statistically significant differences were found in performance between groups we were not powered to reject the null hypothesis. In absolute terms this seemed to be the case for SPAN in our study. However (again in absolute terms) our simple new iSPAN showed relatively better accuracy for non-Caucasian compared to Caucasian patients which suggests that our model at least does not further propagate any racial bias. We hypothesise that this is due to the universal and objective nature of the included features (mTICI and number of passes). This is in keeping with a recent large scale stroke prediction paper which concluded there is:

*“(a) need to expand the pool of risk factors and improve modeling techniques to address observed racial disparities and improve model performance”<sup>21</sup>*

## Conclusion

This study demonstrates the potential of both machine learning models, such as XGB, and the integration of additional variables into traditional clinical scores, like SPAN, to improve outcome predictions for patients with stroke undergoing EVT. The iSPAN model, in particular, offers a valuable alternative to "black box" machine learning models by combining enhanced predictive performance with simplicity, interpretability, and ease of use. These findings underscore the importance of refining prognostic tools to guide clinical decision-making, ultimately improving patient outcomes in acute ischemic stroke care.

## Figures

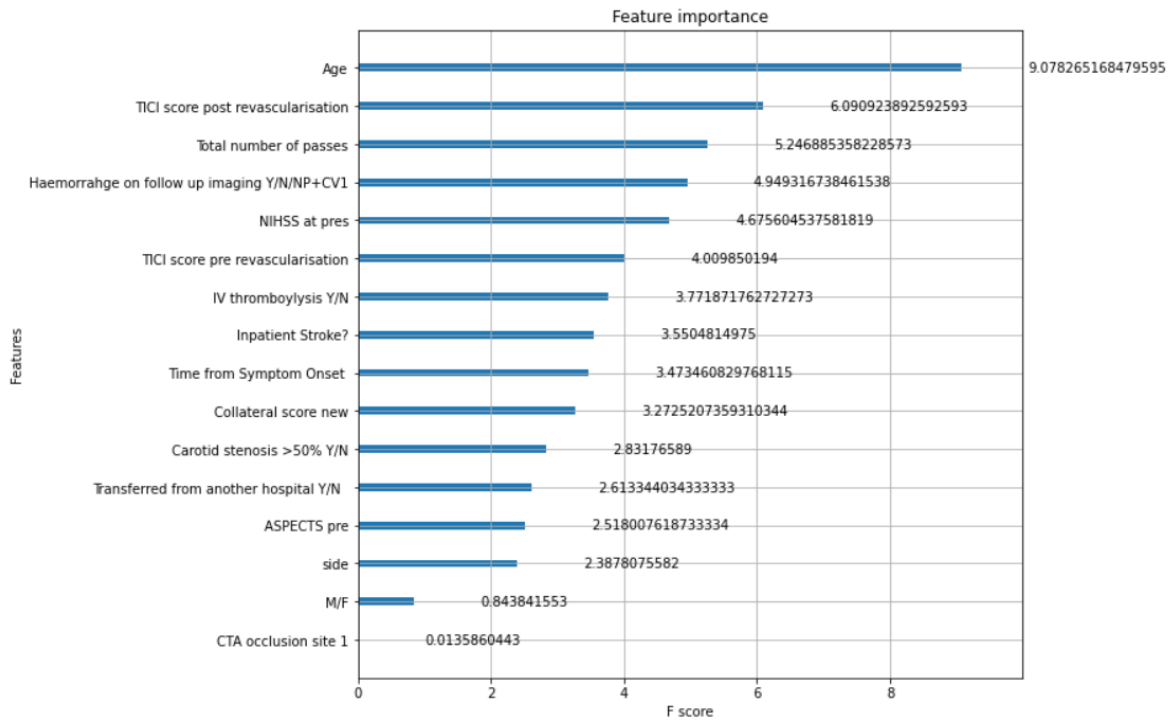


Figure 1

Title: XGB Feature Importance (Training)

Legend: The feature importance with respect to information gain of the best performing XGB model.



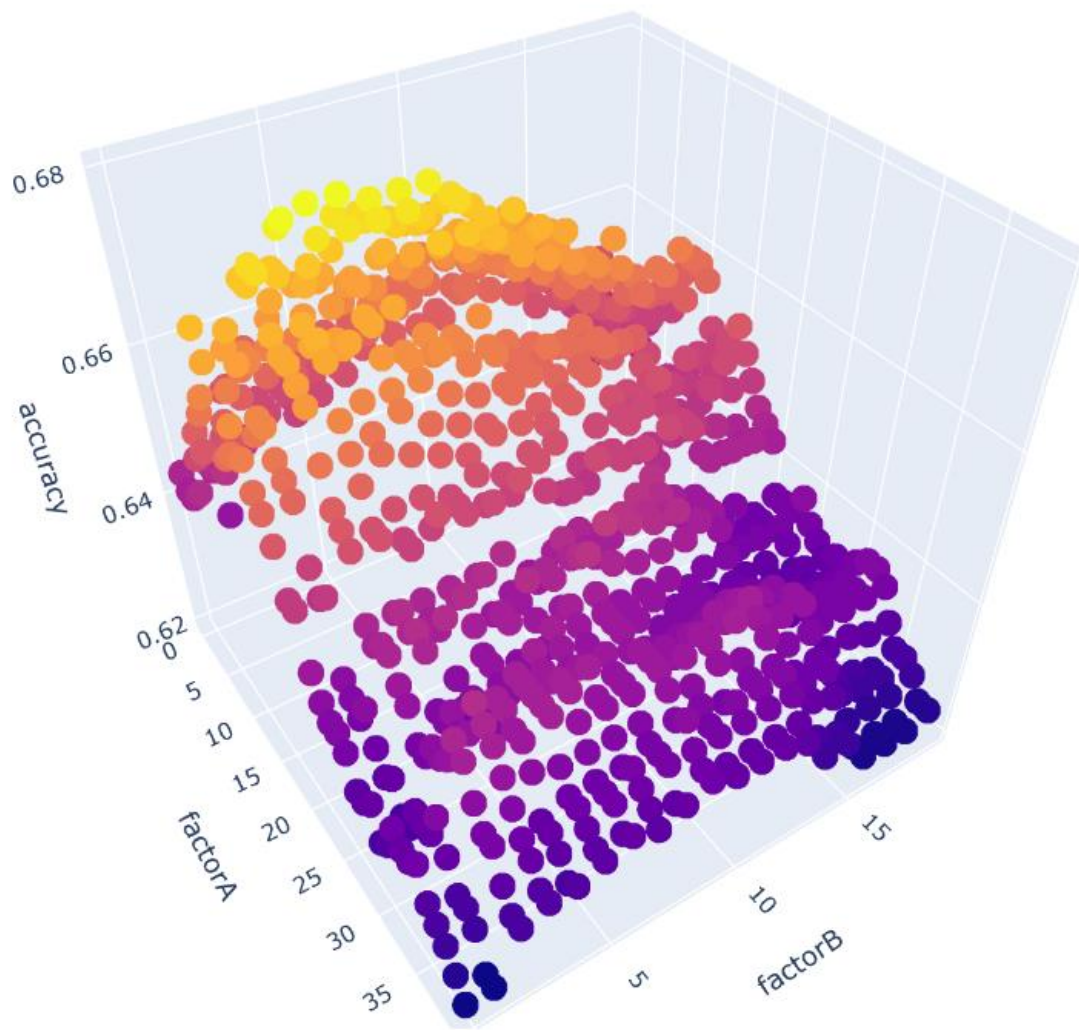


Figure 2:

Title: Optimisation of iSPAN

Legend: Training set weighting optimisation of mTICI (factorA) and number of passes (factorB)

Tables

Table 1

<b>Patient demographic information</b>		
<b>Local Data</b>		
<b>Total Number</b>		812
<b>Sex</b>	Male	415
	Female	397
<b>Mean Age (Mean, Range)</b>		73 (18-93)
<b>NIHSS at presentation</b>		15.3 (2-34)
<b>Number of passes</b>		2.5 (1-16)
<b>mTICI post revascularisation</b>	(Median)	2B (0-3)
<b>Day 90 mRS</b>		2.85 (0-6)
	0, 1 or 2	401
	3 or greater	411
<b>Ethnicity</b>	No data	
<b>External Data</b>		
<b>Total Number</b>		63
<b>Sex</b>	Male	21
	Female	42
<b>Mean Age (Range)</b>		74 (28-92)
<b>NIHSS at presentation</b>		14.6 (2-31)
<b>Number of passes (median)</b>		2.25 (1-12)
<b>mTICI post revascularisation</b>	(Median)	2B (0-3)
<b>Day 90 mRS</b>		2.97 (0-6)
	0, 1 or 2	29
	3 or greater	34
<b>Ethnicity</b>	Caucasian	38
	Asian	10
	African American	6
	Hispanic/Latinx	4
	Middle Eastern	2
	Unknown	2
	Pacific Islands	1

Table 2

Title: Results of different models and clinical scores on test set

<b>Model</b>	<b>Accuracy</b>	<b>Sensitivity</b>	<b>Specificity</b>
MLP	0.612	0.533	0.689
kNN	0.566	0.59	0.541
SVM	0.545	0.933	0.163
XGB	0.738	0.783	0.693
SPAN	0.623	0.967	0.290
iSPAN	0.737	0.817	0.661

Table 3

Title: Results of iSPAN on test and validation sets

		SPAN	iSPAN	
<b>Internal</b>				
	Sensitivity	0.967	0.817	
	Specificity	0.290	0.661	
	Accuracy	0.628	0.737	
<b>External</b>				
	Sensitivity	0.897	0.735	
	Specificity	0.471	0.862	
	Accuracy	0.67	0.79	
<b>Sub Group</b>		SPAN	iSPAN	Relative change (%)
<b>Caucasian</b>				
	Sensitivity	0.833	0.882	0.06
	Specificity	0.654	0.81	0.29
	Accuracy	0.71	0.84	0.18
<b>Other</b>				
	Sensitivity	0.40	0.66	0.4
	Specificity	0.90	0.80	-0.11
	Accuracy	0.60	0.72	0.2

## References

1. Katan M, Luft A. Global Burden of Stroke. *Semin Neurol*. 2018;38:208–211.
2. Saver JL. Time Is Brain—Quantified. *Stroke*. 2006;37:263–266.
3. Almekhlafi MA, Goyal M, Dippel DWJ, Majoie CBLM, Campbell BCV, Muir KW, Demchuk AM, Bracard S, Guillemin F, Jovin TG, et al. Healthy Life-Year Costs of Treatment Speed From Arrival to Endovascular Thrombectomy in Patients With Ischemic Stroke. *Jama Neurol*. 2021;78:709–717.
4. Saver JL, Goyal M, Bonafe A, Diener H-C, Levy EI, Pereira VM, Albers GW, Cognard C, Cohen DJ, Hacke W, et al. Stent-Retriever Thrombectomy after Intravenous t-PA vs. t-PA Alone in Stroke. *New Engl J Medicine*. 2015;372:2285–2295.
5. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, Yavagal DR, Ribo M, Cognard C, Hanel RA, et al. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *New Engl J Medicine*. 2018;378:11–21.
6. (WSO) F the AA of NS (AANS) American Society of Neuroradiology (ASNR), Cardiovascular and Interventional Radiology Society of Europe (CIRSE), Canadian Interventional Radiology Association (CIRA), Congress of Neurological Surgeons (CNS), European Society of Minimally Invasive Neurological Therapy (ESMINT), European Society of Neuroradiology (ESNR), European Stroke Organization (ESO), Society for Cardiovascular Angiography and Interventions (SCAI), Society of Interventional Radiology (SIR), Society of NeuroInterventional Surgery (SNIS), and World Stroke Organization, Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, Eesa M, Fischer U, Hausegger K, et al. Multisociety Consensus Quality Improvement Revised Consensus Statement for Endovascular Therapy of Acute Ischemic Stroke. *Int J Stroke*. 2017;13:612–632.
7. Mistry EA, Mistry AM, Nakawah MO, Chitale RV, James RF, Volpi JJ, Fusco MR. Mechanical Thrombectomy Outcomes With and Without Intravenous Thrombolysis in Stroke Patients. *Stroke*. 2018;48:2450–2456.
8. Duffy S, McCarthy R, Farrell M, Thomas S, Brennan P, Power S, O’Hare A, Morris L, Rainsford E, MacCarthy E, et al. Per-Pass Analysis of Thrombus Composition in Patients With Acute Ischemic Stroke Undergoing Mechanical Thrombectomy. *Stroke*. 2019;50:1156–1163.
9. Rossi R, Fitzgerald S, Molina S, Mereuta OM, Douglas A, Pandit A, Santos AMS, Murphy B, Alderson J, Brennan P, et al. The administration of rtPA before mechanical thrombectomy in acute ischemic stroke patients is associated with a significant reduction of the retrieved clot area but it does not influence revascularization outcome. *J Thromb Thrombolys*. 2021;51:545–551.
10. Bres-Bullrich M, Fridman S, Sposato LA. Relative Effect of Stroke Severity and Age on Outcomes of Mechanical Thrombectomy in Acute Ischemic Stroke. *Stroke*. 2021;52:2846–2848.

11. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, Chihara H, Fukumitsu R, Okawa M, Yamana N, et al. Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning. *Stroke*. 2019;50:2379–2388.
12. Teo YH, Lim ICZY, Tseng FS, Teo YN, Kow CS, Ng ZHC, Ko NCK, Sia C-H, Leow AST, Yeung W, et al. Predicting Clinical Outcomes in Acute Ischemic Stroke Patients Undergoing Endovascular Thrombectomy with Machine Learning. *Clin Neuroradiol*. 2021;31:1121–1130.
13. Wardlaw JM, Murray V, Berge E, Zoppo GJ del. Thrombolysis for acute ischaemic stroke. *Cochrane Db Syst Rev*. 2014;2016:CD000213.
14. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg*. 2020;12:156.
15. Kelly B, Martinez M, Do H, Hayden J, Huang Y, Yedavalli V, Ho C, Keane PA, Killeen R, Lawlor A, et al. DEEP MOVEMENT: Deep learning of movie files for management of endovascular thrombectomy. *Eur. Radiol*. 2023;
16. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *Npj Digital Medicine*. 2018;1:40.
17. Gastouniotti A, Kontos D. Is It Time to Get Rid of Black Boxes and Cultivate Trust in AI? *Radiology Artif Intell*. 2020;2:e200088.
18. Alvarado R. Should we replace radiologists with deep learning? Pigeons, error and trust in medical AI. *Bioethics*. 2022;36:121–133.
19. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30–36.
20. Rouzrokh P, Khosravi B, Faghani S, Moassefi M, Garcia DVV, Singh Y, Zhang K, Conte GM, Erickson BJ. Mitigating Bias in Radiology Machine Learning: 1. Data Handling. *Radiology Artif Intell*. 2022;4:e210290.
21. Hong C, Pencina MJ, Wojdyla DM, Hall JL, Judd SE, Cary M, Engelhard MM, Berchuck S, Xian Y, D'Agostino R, et al. Predictive Accuracy of Stroke Risk Prediction Models Across Black and White Race, Sex, and Age Groups. *Jama*. 2023;329:306–317.
22. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, initiative for the S. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370:1453–1457.
23. Rangaraju S, Aghaebrahim A, Streib C, Sun C-H, Ribo M, Muchada M, Nogueira R, Frankel M, Gupta R, Jadhav A, et al. Pittsburgh Response to Endovascular therapy (PRE) score: optimizing patient selection for endovascular therapy for large vessel occlusion strokes. *J Neurointerv Surg*. 2015;7:783.

24. Saposnik G, Guzik AK, Reeves M, Ovbiagele B, Johnston SC. Stroke Prognostication using Age and NIH Stroke Scale. *Neurology*. 2013;80:21–28.

25. Riley RD, Snell KI, Ensor J, Burke DL, Jr FEH, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statist. Med*. 2019;38:1276–1296.

26. Alawieh A, Zaraket F, Alawieh MB, Chatterjee AR, Spiotta A. Using machine learning to optimize selection of elderly patients for endovascular thrombectomy. *J Neurointerv Surg*. 2019;11:847.

27. Shwartz-Ziv R, Armon A. Tabular data: Deep learning is not all you need. *Inform Fusion*. 2022;81:84–90.

28. Hu Y, Yang T, Zhang J, Wang X, Cui X, Chen N, Zhou J, Jiang F, Zhu J, Zou J. Dynamic Prediction of Mechanical Thrombectomy Outcome for Acute Ischemic Stroke Patients Using Machine Learning. *Brain Sci*. 2022;12:938.

29. Saiegh FA, Munoz A, Velagapudi L, Theofanis T, Suryadevara N, Patel P, Jabre R, Chen C, Shehabeldin M, Gooch MR, et al. Patient and procedure selection for mechanical thrombectomy: Toward personalized medicine and the role of artificial intelligence. *J Neuroimaging*. 2022;32:798–807.

30. Velagapudi L, Mouchtouris N, Schmidt RF, Vuong D, Khanna O, Sweid A, Sadler B, Saiegh FA, Gooch MR, Jabbour P, et al. A Machine Learning Approach to First Pass Reperfusion in Mechanical Thrombectomy: Prediction and Feature Analysis. *J Stroke Cerebrovasc Dis*. 2021;30:105796.

31. Hamann J, Herzog L, Wehrli C, Dobrocky T, Bink A, Piccirelli M, Panos L, Kaesmacher J, Fischer U, Stippich C, et al. Machine-learning-based outcome prediction in stroke patients with middle cerebral artery-M1 occlusions and early thrombectomy. *Eur J Neurol*. 2021;28:1234–1243.

32. Lu S, Zhang J, Wu R, Cao Y, Xu X, Li G, Liu S, Shi H, Wu F. Use of Machine Learning Algorithms to Predict the Outcomes of Mechanical Thrombectomy in Acute Ischemic Stroke Patients With an Extended Therapeutic Time Window. *J Comput Assist Tomo*. 2022;46:775–780.

33. Yao Z, Mao C, Ke Z, Xu Y. An explainable machine learning model for predicting the outcome of ischemic stroke after mechanical thrombectomy. *J Neurointerv Surg*. 2022;jnis-2022-019598.

34. Tariq MB, Jeevarajan J, Nahhas MI, Salazar-Marioni S, Abdelkhaleq R, Niktabe A, Sheth S. Abstract WMP93: Additional Endovascular Thrombectomy Passes After An Initial TIC12b Reperfusion To Improve Final Reperfusion Grade Does Not Improve Functional Outcomes. *Stroke*. 2023;54:AWMP93–AWMP93.

35. Kelly B, Kirwan A, Quinn M, Kelly AM, Mathur P, Lawlor A, Killeen R. The ethical matrix as a method for Involving People living with disease and the wider Public (PPI) in near-term artificial intelligence research. *Radiography*. 2023;

36. Daye D, Wiggins WF, Lungren MP, Alkasab T, Kottler N, Allen B, Roth CJ, Bizzo BC, Durniak K, Brink JA, et al. Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? *Radiology*. 2022;305:555–563.

37. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digital Heal*. 2021;3:e195–e203.