AJNR

Comprehensive segmentation of gray matter structures on T1-weighted brain MRI: A Comparative Study of CNN, CNN hybrid-transformer or -Mamba architectures

Yujia Wei, Jaidip Manikrao Jagtap, Yashbir Singh, Bardia Khosravi, Jason Cai, Jeffrey L Gunter and Bradley J Erickson

AJNR Am J Neuroradiol published online 21 October 2024
http://www.ajnr.org/content/early/2024/10/21/ajnr.A8544

Published October 21, 2024 as 10.3174/ajnr.A8544

ORIGINAL RESEARCH

Comprehensive segmentation of gray matter structures on T1-weighted brain MRI: A Comparative Study of CNN, CNN hybrid-transformer or -Mamba architectures

Yujia Wei, Jaidip Manikrao Jagtap, Yashbir Singh, Bardia Khosravi, Jason Cai, Jeffrey L Gunter, Bradley J Erickson*

ABSTRACT

BACKGROUND AND PURPOSE: Recent advances in deep learning have shown promising results in medical image analysis and segmentation. However, most brain MRI segmentation models are limited by the size of their datasets and/or the number of structures they can identify. This study evaluates the performance of six advanced deep learning models in segmenting 122 brain structures from T1-weighted MRI scans, aiming to identify the most effective model for clinical and research applications.

MATERIALS AND METHODS: 1,510 T1-weighted MRIs were used to compare six deep-learning models for the segmentation of 122 distinct gray matter structures: nnU-Net, SegResNet, SwinUNETR, UNETR, U-Mamba_BOT and U-Mamba_ Enc. Each model was rigorously tested for accuracy using the Dice Similarity Coefficient (DSC) and the 95th percentile Hausdorff Distance (HD95). Additionally, the volume of each structure was calculated and compared between normal control (NC) and Alzheimer's Disease (AD) patients.

RESULTS: U-Mamba_Bot achieved the highest performance with a median DSC of 0.9112 [IQR:0.8957, 0.9250]. nnU-Net achieved a median DSC of 0.9027 [IQR: 0.8847, 0.9205] and had the highest HD95 of 1.392[IQR: 1.174, 2.029]. The value of each HD95 (<3mm) indicates its superior capability in capturing detailed brain structures accurately. Following segmentation, volume calculations were performed, and the resultant volumes of normal controls and AD patients were compared. The volume changes observed in thirteen brain substructures were all consistent with those reported in existing literature, reinforcing the reliability of the segmentation outputs.

CONCLUSIONS: This study underscores the efficacy of U-Mamba_Bot as a robust tool for detailed brain structure segmentation in T1-weighted MRI scans. The congruence of our volumetric analysis with the literature further validates the potential of advanced deep-learning models to enhance the understanding of neurodegenerative diseases such as AD. Future research should consider larger datasets to validate these findings further and explore the applicability of these models in other neurological conditions.

ABBREVIATIONS: AD = Alzheimer's Disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; DSC = Dice Similarity Coefficient; HD95 = the 95th Percentile Hausdorff Distance; IQR = Interquartile Range; NC = Normal Control; SSMs = State-space Sequence Models.

Received month day, year; accepted after revision month day, year.
From the Department of Radiology, Mayo Clinic, Rochester, MN (Y.W., J.M.J., Y.S., B.K., J.C., J.L.G., B.J.E.).

The authors declare no conflicts of interest related to the content of this article.

Please address correspondence to Bradley J Erickson, MD, PhD., Department of Radiology, Mayo Clinic, 200 1st St. SW, Rochester, MN, 55905; bje@mayo.edu

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies have demonstrated the utility of convolutional neural networks (CNNs) and hybrid transformer models in medical image segmentation, particularly in neuroimaging. U-Net-based architectures have been widely adopted for their ability to capture spatial details, while transformer models show promise in capturing global dependencies. However, many approaches still face limitations, such as computational resource demands and high memory usage, when applied to large-scale datasets. The introduction of structured state-space models (SSMs), such as U-Mamba, provides a new perspective on improving both segmentation accuracy and computational efficiency in biomedical imaging.

KEY FINDINGS: Our study found that U-Mamba_Bot outperformed other models, achieving the highest Dice Similarity Coefficient (DSC) of 0.9112. It also exhibited competitive training and inference times compared to other architectures.

KNOWLEDGE ADVANCEMENT: The U-Mamba model's integration of structured state-space mechanisms addresses some of the limitations of traditional CNN and transformer-based models, particularly in capturing long-range dependencies with lower computational cost. These findings highlight U-Mamba's potential for enhancing neuroimaging analysis in clinical applications.

Copyright 2024 by American Society of Neuroradiology.

INTRODUCTION

MRI can deliver superior spatial and contrast resolution and has become a cornerstone in diagnosing and treating neurological disease. Instance segmentation refers to delineating intracranial structures (segmentation) and assigning individual labels to every structure, which is essential in studying brain MRI. It provides valuable information for structural analysis, volumetric assessment, surgical planning, and image-guided intervention. For example, several studies from the Alzheimer's Disease Neuroimaging Initiative (ADNI) demonstrated that alterations in intracranial structural volumes (as quantified using segmentation tools) correlate with outcome measures in clinical trials for AD. A study from Radue et al. demonstrated that brain volume loss correlated with clinical and radiological outcomes in multiple sclerosis patients[1], and Mora et al. demonstrated that patients with medial temporal lobe epilepsy exhibit a consistent pattern of gray matter atrophy on MRI[2], suggesting that a common pathophysiological process may be responsible for the disease. Thus, diagnosing neurological and neuropsychiatric diseases necessitates a comprehensive understanding of subcortical structures, and it is crucial to grasp both the structural and functional characteristics of the brain.

Automated segmentation methods have been developed to differentiate between frontotemporal dementia and AD based on MRI images [3]. Recent advances in deep learning have shown promising results in medical image analysis and segmentation[4]. State-of-the-art methods, such as U-Net and transformers, have achieved significant success in medical image segmentation [5]. The CNN-based U-Net architecture has been widely utilized in various fields, particularly medical image segmentation, due to its effectiveness in capturing spatial information and features. Though initially developed for natural language processing tasks, transformers have recently been adapted for medical image segmentation and offer promising results[6,7]. Although U-Net can better recognize local features and transformers can capture global features, given their respective advantages and disadvantages, many current approaches are combining the two, for example, UNETR[8], SwinUNETR[9] for more accurate results and more robust performance. However, certain shortcomings still exist, such as being resource-intensive, and having high memory and computational requirements.

Recently, state-space sequence models (SSMs), particularly structured state-space sequence models, have emerged as efficient and powerful components for constructing deep networks that deliver top-tier performance in continuous long-sequence data analysis[10]. Mamba improved the structured state-space sequence models model by introducing a selective mechanism that allows the model to select relevant information depending on the input[11]. U-Mamba was newly developed for general-purpose biomedical image segmentation with a self-adapting function network based on the innovative hybrid CNN-SSM architecture. Compared to the currently popular deep learning network architectures, nnU-Net, SegResNet[12], and transformer-based SwinUNETR[9], U-Mamba achieved the best results in image segmentation for abdominal MRI, instruments in endoscopy, and cells in microscopy[13].

In this study, we compared the newly released U-Mamba automatic segmentation model and previous state-of-the-art segmentation models for whole brain substructure segmentation on T1-weighted MRI. We utilized the ADNI database to conduct comparative analyses of the most popular current models for medical image segmentation for almost all cortical subregions and nuclei of the human brain. This sets the stage for further understanding the relationship between brain structural changes and diseases, uncovering unknown disease mechanisms, and providing highly automated and robust tools. We believe these have significant potential for clinical application.

MATERIALS AND METHODS

Data collection

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

Table 1: Demographics of the dataset we obtained for the study

Diagnosis	Number of cases (n=1510)	Age (years) Mean (SD) or Median [IQR]	Sex (Male/Female)
NC	359	73.6 [70.8, 78.1]	176/183
AD	304	75.8 [70.9, 80.6]	166/138
EMCI	270	71.1 (7.4)	145/125
LMCI	141	72.2 (7.5)	75/66
MCI	338	75.2 [70.4, 80.3]	222/116
SMC	98	71.2 [67.5, 76.5]	56/42

* Normal Controls (NC), Alzheimer's Disease (AD), and various stages of Mild Cognitive Impairment (MCI), including Early MCI (EMCI), Late MCI (LMCI), and Significant Memory Concern (SMC).

This study analyzed 20,056 randomly selected T1-weighted MRIs from the ADNI datasets. Since each patient typically undergoes multiple MRI scans, we retained the initial MRI from each patient in the ADNI database. We excluded cases lacking detailed information and ultimately obtained 1,510 MRI scans for this study (Table 1). Normally distributed data are expressed as means (SD), and non-normally distributed data are expressed as medians with interquartile range (median [IQR:25th, 75th]).

Data labeling and Pre-processing

To generate ground truth labels (pixel-level segmentation masks), we linearly registered the Mayo Clinic Adult Lifespan Template and Atlas to every volume using ANTs tool (https://github.com/ANTsX/ANTs) resulting in 122 gray matter structures[14, 34]. Board-certified radiologists visually inspected all volumes to ensure good image quality and proper registration. We divided the dataset into 768 training, 192 validations, and 550 for testing (all from different patients) for model evaluation. Table 1 summarizes the demographics and characteristics of the patients in this dataset.

Since our study is focused on brain segmentation, the deep-learning model requires the removal of non-brain tissue to increase the model's performance. A Deep learning-based brain extraction tool, HD-BET[15], was used on T1-weighted MR images.

Model training

All of the deep learning models were adapted to work within the nnU-Net framework. nnU-Net is a deep learning-based segmentation method that automatically configures and runs the entire segmentation pipeline for any biomedical image dataset, including preprocessing, data augmentation, model training, and post-processing. The pipeline handles hyperparameter tuning and does not require any changes to the network architecture. Therefore, it provides a perfect environment for comparing U-Mamba with other methods. Also, it enables U-Mamba to be easily adapted to a wide range of segmentation tasks. In the nnU-Net framework, the patch size is 128x128x128 with a batch size equal to two. The Adam optimizer with an initial learning rate of 0.01 was used to optimize network weights, with the momentum set at 0.99. An empirical combination of Dice loss with cross-entropy loss in nnU-Net has enhanced training stability and improved segmentation accuracy[13,16]. To ensure a fair comparison, we also implemented SegResNet[12], UNETR[8], and SwinUNETR[9] into the nnU-Net framework and utilized nnU-Net recommended and default optimizers for model training[13]. For more detailed information on how each model was adapted within the nnU-Net framework, please refer to the GitHub repository: https://github.com/wyjzll/U-Mamba. This repository includes comprehensive documentation and code examples illustrating the integration process. To ensure the reproducibility of this study and maintain consistency in the code version used, independent of updates by the original U-Mamba authors[13], we forked the original U-Mamba GitHub repository (https://github.com/bowang-lab/U-Mamba). This approach guarantees that the code remains unchanged, allowing others to reliably replicate our findings. All models were trained on one GPU (NVIDIA A100 80G SXM) for 1000 epochs with random initial weights. The entire process of this study is presented in Figure 1.

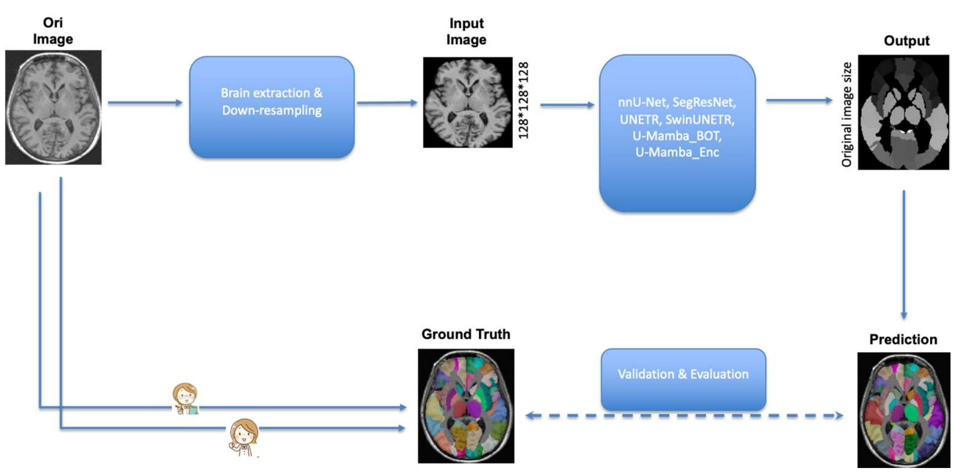

FIG 1. An overview of the study flow. (a) T1-weighted MRI images were first processed through a brain extraction step using the Brain Extraction Tool-HD-BET. (b) The processed image then serves as the input to each of the deep learning models, which are responsible for segmenting the brain into different anatomical regions. (c) The evaluation process was conducted by comparing the segmentation results for each model with the ground truth.

Evaluation

The performance of the models in segmenting all 122 brain substructures was assessed using the DSC and the 95th percentile Hausdorff Distance (HD95).

The DSC is defined in equation (1):

$$DSC = \frac{2\,|X \cap Y|}{|X| + |Y|} \qquad (1)$$

Here, X and Y represent the ground truth segmentation and model segmentation. The $|\cdot|$ indicates the number of elements in the set and $\cap$ represents the intersection of the sets.

The HD is defined in equation (2):

$H(A, B) = max(h(A, B), h(B, A))$ (2)

where:

A and B are the two sets of points (e.g., the edges of the segmented regions for the ground truth and for a model)

h(A,B) is defined as maxa ∈ A minb ∈ B d(a, b)

h(B,A) is defined as maxb ∈B mina ∈A d(b, a)

d(a, b) is the distance between points a and b (typically Euclidean distance).

The 95th percentile Hausdorff Distance is calculated by taking the 95th percentile of all the computed distances rather than the maximum, helping to ignore the most extreme values that might be due to noise or other anomalies. This makes it a robust measure for assessing the accuracy of medical image segmentations, where outliers can skew the results.

Statistics

We use the Kolmogorov-Smirnov test for all normality tests for all of the groups. Two-sample t-tests were used to compare differences in normally distributed data between groups. If data were not normally distributed, the data were tested using nonparametric approaches (Mann-Whitney test). The statistical tests were performed using GraphPad Prism 10.0 and SciPy 1.8.0. Results with $p<0.05$ were indicated by asterisks. d

This article follows the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis Checklist.[35]

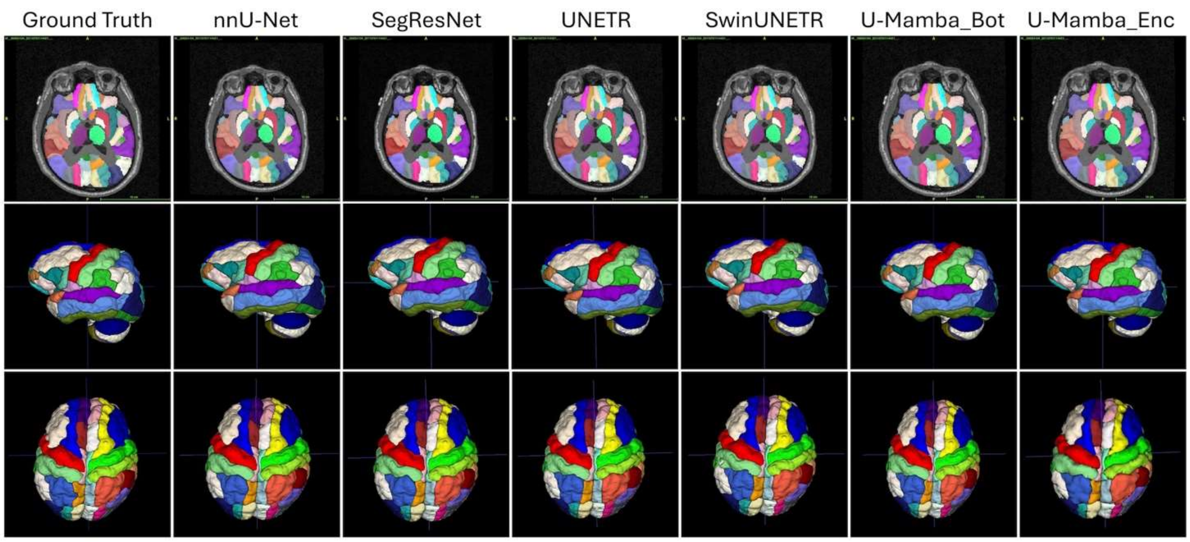

FIG 2. Representative images generated by the deep-learning models vs. human-labeled Ground Truth.

RESULTS

Model performance

A representative mask generated by our model is shown in Figure 2. All 122 gray matter regions were segmented successfully in all of the models (that is, no structure had a DSC of 0 for any model, Supplemental table 1 & 2). We have provided downloadable result samples on our GitHub page (https://github.com/wyjzll/Brain_Segmentation) for interested readers to further reference. The models evaluated were nnU-Net, SegResNet, UNETR, SwinUNETR, U-Mamba_BOT, and U-Mamba_Enc. Each model's effectiveness is assessed using the DSC and HD95. A success rates of DSC >0.9 or HD95 <3 mm was defined to ensure high accuracy and reliability in clinical settings (Table 2). Among these models, U-Mamba_Bot showed superior segmentation accuracy with the highest DSC value of 0.9112 [IQR: 0.8957, 0.9250], achieving high success rates of 68.85% for DSC>0.9, this model, however, did not have the best HD95 score, suggesting a potential trade-off between general overlap and boundary precision. U-Mamba_Enc demonstrated competitive performance, closely matching nnU-Net in DSC (0.8968 [IQR:0.8801, 0.9155]), but with a higher HD95 of 1.544 [IQR: 1.224, 2.318]. In contrast, the SegResNet, while having a competitive DSC score (0.9033), exhibited a median HD95 (1.449 mm), which could imply higher boundary accuracy compared to other models. The UNETR model, as presented in the chart, showed a DSC of 0.8709 with an interquartile range (IQR) from 0.8521 to 0.8978, which was the lowest among the models tested. This suggests that it may have some limitations in achieving a high degree of overlap between the predicted and true segmentations compared to the other models.

Table 2: Results summary of trained segmentation models on T1 weighted MRI datasets.

Model	DSC, Median (IQR)	HD95(mm), Median (IQR)	Success Rate (DSC >0.9)	Success Rate (HD95<3mm)
nnU-Net V2	0.9027 (0.8847, 0.9205)	1.392 (1.174, 2.029)	52.46%	93.44%
SegResNet	0.9033 (0.8846, 0.9185)	1.449 (1.158, 2.092)	53.28%	93.44%
UNETR	0.8709 (0.8521, 0.8978)	2.053 (1.451, 2.949)	21.31%	76.23%
SwinUNETR	0.8944 (0.8784, 0.9123)	1.698 (1.288, 2.563)	41.80%	88.52%
U-Mamba_Bot	0.9112 (0.8957, 0.9250)	2.699 (2.196, 4.039)	68.85%	60.66%
U-Mamba_Enc	0.8968 (0.8801, 0.9155)	1.544 (1.224, 2.318)	45.90%	90.98%

* U-Mamba_Bot: only use the U-Mamba block in the bottleneck. U-Mamba_Enc: all encoder blocks are U-Mamba blocks.

As shown in Table 3, the training times varied significantly among models. nnU-Net V2 has the longest epoch time at 412 seconds, indicating that it may require considerable computational resources and time for training. U-Mamba_Enc had the fastest training time, completing an epoch in just 169 seconds. SegResNet and UNETR also showed relatively quick epoch times at 183 and 191 seconds, respectively, making them suitable for environments where faster model training is beneficial. SwinUNETR and U-Mamba_Bot fell in the middle, with epoch times of 314 and 273 seconds, respectively, balancing computational complexity and training speed.

Table 3: A comparison of the epoch training times and inference time per image for different deep learning models. The times listed are in seconds (s), performed on a NVidia A100 80G SXM GPU.

	nnU-Net	SegResNet	UNETR	SwinUNETR	U-Mamba_Bot	U-Mamba_Enc
Training Time per Epoch (s)	412	183	191	314	273	169
Inference Time per Image(s)	6.3	3.8	12.1	3.7	3.6	6.7

To assess the effectiveness of the models, we calculated the degree of correspondence between the volumes of individual brain regions in each model's predictions and the ground truth. Figure 3 shows the performance of various deep learning models in segmenting brain substructures. Each model's accuracy was evaluated based on its ability to match ground truth measurements, with the results categorized into significant differences ($p<0.05$) and non-significant differences ($p>0.05$). The models included in the analysis are nnU-Net, SegResNet, UNETR, SwinUNETR, U-Mamba_Bot, and U-Mamba_Enc. SegResNet and U-Mamba_Bot aligned more closely to the ground truth, indicated by more non-significant differences versus ground truth (117 and 118, respectively). nnU-Net demonstrated a high discrepancy from ground truth with 122 significant differences.

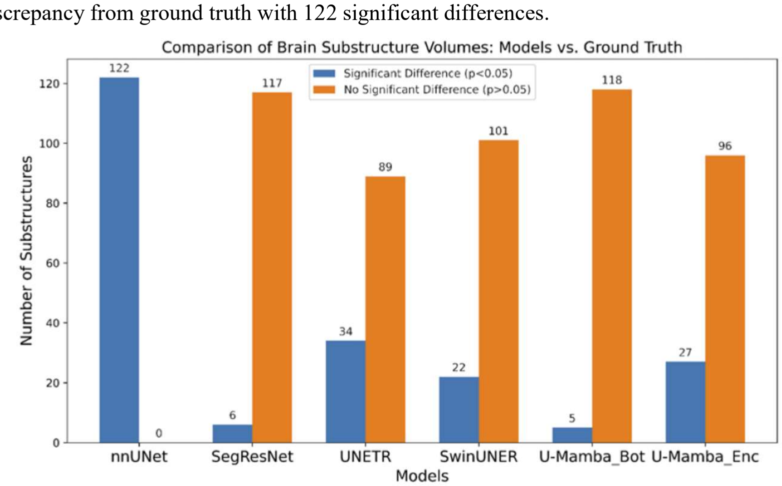

FIG 3. Comparison of the number of brain substructure volumes that had significant differences from ground truth using different models.

Clinical testing

Patients with AD are intricately linked to changes in brain structure and volume characterized by brain atrophy. Understanding these alterations through neuroimaging studies is crucial for the early detection and monitoring of disease progression. In this part, we analyzed the prediction results produced by U-Mamba_Bot on the test set, which includes 134 healthy individuals and 112 AD patients. We found that among all the brain areas that atrophied in AD patients, the Amygdala exhibited the most significant volume reduction compared to the NC group, with the left and right sides shrinking by 13.03% and 10.03%, respectively. This was followed by the bilateral entorhinal cortex, which decreased by 8.60% on the left side and 9.33% on the right. In contrast, the volumes of the Caudate in the NC and AD groups increased by 8.74% and 7.27%, respectively, which aligns with the findings reported in the literature[17]. For ease of presentation, we only show the volume ranges of the 13 brain regions that have statistically significant differences between the NC and AD groups (Figure 4), based on the image segmentation results generated by the U-Mamba_Bot model.

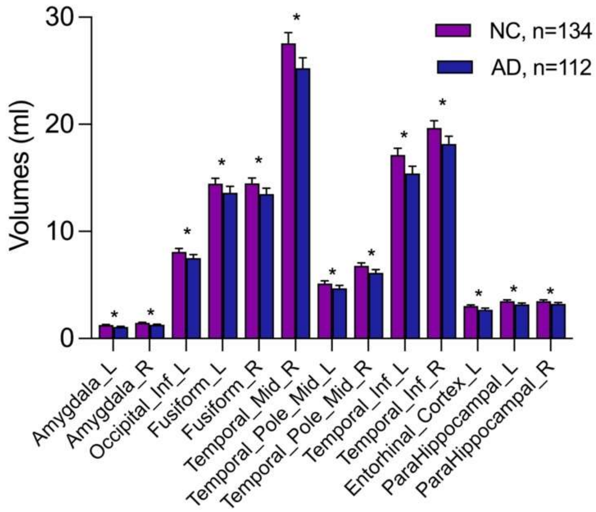

FIG 4. Comparison of Brain Substructure Volumes between NC and AD. The dual-colored bars represent quantified volumes for each respective brain substructure, highlighting significant discrepancies between the two groups (* p<0.05, Mann-Whitney test).

DISCUSSION

In this paper, we have chosen the latest U-Mamba model to conduct what we believe to be the most extensive sub-structural segmentation of the brain to date. Our comprehensive experiment effectively highlights the variability in the accuracy of most deep learning models in replicating precise brain substructure volumes, providing insights into their reliability for clinical and research applications, and suggested model selection based on specific requirements for accuracy and margin precision in future clinical applications. The results indicate that U-Mamba's performance surpasses that of existing CNN and Transformer-based segmentation networks across different modes and segmentation targets. In particular, U-Mamba has significantly faster training speeds compared to CNN and Transformer architectures. This is instrumental in addressing challenges posed by the local nature of CNNs and the computational complexity of Transformers, which affect long-range modeling. This advantage is largely attributed to the architectural design of U-Mamba, which is capable of extracting multi-scale local features while capturing long-range dependencies.

The nnU-Net framework, based on the U-Net architecture, has demonstrated exceptional performance in various segmentation tasks, surpassing state-of-the-art models in international biomedical image segmentation challenges. Its success can also be attributed to adaptive preprocessing, extensive data augmentation, model ensembling, and aggregating tiled predictions, which collectively contribute to its consistently high performance across a wide range of tasks16,18. nnU-Net can be configured to execute the entire segmentation pipeline automatically; In addition, it offers a range of features that make it highly adaptable and effective across different models and tasks. No other specific data preprocessing (beyond brain extraction) is needed in this part16. We have chosen the nnU-Net as our segmentation network backbone, which enabled us to focus on implementing the network while managing other variables like image preprocessing and data augmentation. This arrangement facilitates a fair comparison of U-Mamba with other methods under consistent conditions, where the network architecture is the sole variable that differs.

In evaluating the efficiency of different deep learning models, it's essential to consider the training times, which can vary significantly among models. Our results demonstrated that U-Mamba_Enc has the fastest training time among the most popular models. However, U-Mamba_Bot showed the advantage of having the quickest inference time among all the models we evaluated. Based on Gu et al. and Ma et al. 's study, Mamba advances SSMs in discrete data modeling, such as text and genomes, through two significant enhancements. Firstly, Mamba introduces an input-dependent selection mechanism, a departure from the traditional time- and input-invariant SSMs, enabling effective filtration of information from inputs. This mechanism is achieved by parameterizing the SSM parameters according to the input data11,13. Secondly, Mamba incorporates a hardware-aware algorithm that scales linearly with sequence length and computes the model recurrently with a scan, thereby enhancing processing speed on modern hardware.

The Mamba architecture, which combines SSM blocks with linear layers, is notably simpler and has achieved state-of-the-art performance in various long-sequence domains such as language and genomics. This simplicity translates into significant computational efficiencies in the training and inference phases11,13. Wu et al. explored the core features of Mamba and conceptually determined that it is best suited for tasks involving long sequences and autoregressive features. For vision tasks that don't have these characteristics, such as image classification, they argue that Mamba may not be necessary. However, while detection and segmentation tasks are not autoregressive, they do involve long sequences19. Interestingly, our study also shows that the overall performance of U-Mamba_Bot with U-Mamba block applied only at bottleneck achieves the highest DSC value among all models, exceeding that of U-Mamba_Enc with U-Mamba block applied in all encoder parts. This suggests that replacing all encoder modules with SSM blocks may not necessarily yield optimal accuracy. Therefore, it is worth investigating the potential benefits of applying Mamba to such tasks.

Understanding regional brain volume is crucial for comprehending the pathophysiology of various brain-related diseases. Several studies have investigated the relationship between brain volume and different health issues such as Huntington's disease20, Atrial fibrillation21, Alzheimer's disease, critical illnesses22, multiple sclerosis23, Parkinson's disease24, and migraine25. This study utilized the ADNI database and successfully completed whole-brain substructure image segmentation, followed by the calculation of brain substructure volumes and a comparative analysis with the NC group. We identified 13 brain functional areas (see Figure 4) with significant changes in brain volume, most of which align with findings reported in the literature17,26–32. However, our results did not show the significant reduction in hippocampal volume that other studies have reported33, and this discrepancy merits further investigation. In future studies, we also plan to apply our study to a larger database and different diseases to offer valuable insights for the diagnosis, prognosis, and monitoring of treatment in different neurological conditions.

Limitations

We have found that certain types of intracranial lesions, such as displacement, space-occupying lesions, inflammation, trauma, edema, hemorrhage, etc., can significantly impact the performance of our model. Since the ADNI database primarily consists of imaging data from elderly individuals, including healthy controls and those with various cognitive impairments, even though these subjects may exhibit certain pathological or age-related changes in brain structure, their basic structure remains relatively unchanged. Therefore, our model is not optimized for segmentation tasks involving such structural alterations.

Moreover, we have identified some minor hand-labeling mistakes during the process. For example, cerebrospinal fluid signals adjacent to gyri were occasionally mislabeled as gyri. Although these errors did not significantly affect larger brain regions, they could lead to substantial calculation errors in small brain structures. This is an area that requires improvement for future work.

CONCLUSIONS

In conclusion, we successfully segmented gray matter regions using several popular deep-learning models, including U-Mamba, a newly developed deep-learning architecture. Our extensive experimental findings demonstrate that U-Mamba's performance matches that of existing CNN and Transformer-based segmentation networks across various modes and segmentation targets. Specifically, U-Mamba_Bot exhibits a marked increase in accuracy for segmentation models compared to CNN and Transformer architectures. Furthermore, the variability across different brain regions captured in this study not only reinforces the heterogeneity of Alzheimer's pathology but also may guide targeted research into the development of diagnostic markers and therapeutic strategies focused on the most affected areas. We believe it may be used clinically for feature extraction, morphological analysis, and downstream diagnostic tools, thereby contributing to the development of automated diagnostic and treatment assessment systems.

ACKNOWLEDGMENTS

We acknowledge all the authors of the employed public datasets, allowing the community to use these valuable resources for research purposes. We also thank the authors of nnU-Net (https://github.com/MIC-DKFZ/nnUNet) and U-Mamba (https://github.com/bowang-lab/U-Mamba) for making their valuable code publicly available.

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative is funded by the National Institute on Aging (National Institutes of Health Grant U19 AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health including generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research &Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

REFERENCES

1. Radue E-W, Barkhof F, Kappos L, et al. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. *Neurology* 2015;84:784–93.

2. Mora F, Segovia G, del Arco A. Aging, plasticity and environmental enrichment: structural changes and neurotransmitter dynamics in several areas of the brain. *Brain Res Rev* 2007;55:78–88.

3. Yu Q, Mai Y, Ruan Y, et al. An MRI-based strategy for differentiation of frontotemporal dementia and Alzheimer's disease. *Alzheimer's Research & Therapy* 2021;13:23.

4. Adegun AA, Viriri S, Ogundokun RO. Deep Learning Approach for Medical Image Analysis. *Computational Intelligence and Neuroscience* 2021;2021:e6215281.

5. Chen S, Qiu C, Yang W, et al. Multiresolution Aggregation Transformer UNet Based on Multiscale Input and Coordinate Attention for Medical Image Segmentation. *Sensors* 2022;22:3820.

6. Strudel R, Garcia R, Laptev I, et al. Segmenter: Transformer for Semantic Segmentation. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE; 2021:7242–52.

7. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2021 Jun 3. [Epub ahead of print].

8. Hatamizadeh A, Tang Y, Nath V, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE; 2022:1748–58.

9. Hatamizadeh A, Nath V, Tang Y, et al. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi A, Bakas S, eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing; 2022:272–84.

10. Gu A, Goel K, Ré C. Efficiently Modeling Long Sequences with Structured State Spaces. https://doi.org/10.48550/arXiv.2111.00396.

11. Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. https://doi.org/10.48550/arXiv.2312.00752.

12. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi A, Bakas S, Kuijf H, et al., eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing; 2019:311–20.

13. Ma J, Li F, Wang B. U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. 2024 Jan 9. [Epub ahead of print].

14. Avants BB, Tustison NJ, Stauffer M, et al. The Insight ToolKit image registration framework. *Front Neuroinform* 2014;8.

15. Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping* 2019;40:4952–64.

16. Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–11.

17. Persson K, Bohbot VD, Bogdanovic N, et al. Finding of increased caudate nucleus in patients with Alzheimer's disease. *Acta Neurol Scand* 2018;137:224–32.

18. Kahraman AT, Fröding T, Toumpanakis D, et al. Enhanced Classification Performance using Deep Learning Based Segmentation for Pulmonary Embolism Detection in CT Angiography. https://doi.org/10.1101/2023.04.21.23288861.

19. Yu W, Wang X. MambaOut: Do We Really Need Mamba for Vision? https://doi.org/10.48550/arXiv.2405.07992.

20. Kim R, Müller S, Garcia TP. svReg: Structural varying-coefficient regression to differentiate how regional brain atrophy affects motor impairment for Huntington disease severity groups. *Biometrical Journal* 2021;63:1254–71.

21. Moazzami K, Shao IY, Chen LY, et al. Atrial Fibrillation, Brain Volumes, and Subclinical Cerebrovascular Disease (from the Atherosclerosis Risk in Communities Neurocognitive Study [ARIC-NCS]). *The American Journal of Cardiology* 2020;125:222–8.

22. Walker KA, Gottesman RF, Wu A, et al. Association of Hospitalization, Critical Illness, and Infection with Brain Structure in Older Adults. *Journal of the American Geriatrics Society* 2018;66:1919–26.

23. Rovaris M, Inglese M, van Schijndel RA, et al. Sensitivity and reproducibility of volume change measurements of different brain portions on magnetic resonance imaging in patients with multiple sclerosis. *J Neurol* 2000;247:960–5.

24. He H, Liang L, Tang T, et al. Progressive brain changes in Parkinson's disease: A meta-analysis of structural magnetic resonance imaging studies. *Brain Res* 2020;1740:146847.

25. Chen X-Y, Chen Z-Y, Dong Z, et al. Regional volume changes of the brain in migraine chronification. *Neural Regen Res* 2020;15:1701–8.

26. Poulin SP, Dautoff R, Morris JC, et al. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res* 2011;194:7–13.

27. Holroyd S, Shepherd ML, Downs JH. Occipital atrophy is associated with visual hallucinations in Alzheimer's disease. *J Neuropsychiatry Clin Neurosci* 2000;12:25–8.

28. Convit A, De Leon MJ, Tarshish C, et al. Specific Hippocampal Volume Reductions in Individuals at Risk for Alzheimer's Disease. *Neurobiology of Aging* 1997;18:131–8.

29. Foundas AL, Leonard CM, Mahoney SM, et al. Atrophy of the Hippocampus, Parietal Cortex, and Insula in Alzheimer's Disease: A Volumetric Magnetic Resonance Imaging Study. *Cognitive and Behavioral Neurology* 1997;10:81.

30. Visser P, Verhey F, Hofman P, et al. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *J Neurol Neurosurg Psychiatry* 2002;72:491–7.

31. Devanand DP, Pradhaban G, Liu X, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology* 2007;68:828–36.

32. Köhler S, Black SE, Sinden M, et al. Memory impairments associated with hippocampal versus parahippocampal-gyrus atrophy: an MR volumetry study in Alzheimer's disease. *Neuropsychologia* 1998;36:901–14.

33. Schuff N, Woerner N, Boreta L, et al. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain* 2009;132:1067–77.

34. Schwarz CG, Gunter JL, Wiste HJ, et al. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. Neuroimage Clin 2016;11:802–12.

35. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med 2015;162:55–63.

SUPPLEMENTAL FILES

Supplemental Table 1: Dice Similarity Coefficient of the Test Dataset

Classes#	Name of Substructures	nnU-Net V2	SegResNet	UNETR	SwinUNETR	U-Mamba_Bot	U-Mamba_Enc
1	Precentral_L	0.92599269	0.92215681	0.906049592	0.91239291	0.924567393	0.915722765
2	Precentral_R	0.92896768	0.92573323	0.907375913	0.913173685	0.927749776	0.917681322
3	Frontal_Sup_L	0.87371989	0.86564457	0.83961016	0.8499757	0.869054051	0.852173425
4	Frontal_Sup_R	0.87185391	0.86365074	0.836207137	0.844567259	0.869366745	0.850807363
5	Frontal_Sup_Orb_L	0.88817643	0.87872166	0.842684544	0.87061567	0.891873005	0.873789309
6	Frontal_Sup_Orb_R	0.88589453	0.8771023	0.842029271	0.867409389	0.888982967	0.871105135
7	Frontal_Mid_L	0.92021562	0.914375	0.89595759	0.902436784	0.917545006	0.905988289
8	Frontal_Mid_R	0.91490416	0.9084158	0.883618023	0.893427819	0.911208777	0.897525554
9	Frontal_Mid_Orb_L	0.88884477	0.88292301	0.860929664	0.870531711	0.889639368	0.874012051
10	Frontal_Mid_Orb_R	0.89980957	0.89219074	0.859802107	0.879468022	0.900395222	0.882734028
11	Frontal_Inf_Oper_L	0.89518783	0.89005478	0.865737305	0.881072008	0.899262834	0.887182487
12	Frontal_Inf_Oper_R	0.88370655	0.87639334	0.845690747	0.867240816	0.887291311	0.873807519
13	Frontal_Inf_Tri_L	0.89119353	0.88718536	0.863276604	0.875282538	0.892376392	0.883588806
14	Frontal_Inf_Tri_R	0.88632799	0.88048928	0.853949268	0.867671888	0.887271704	0.875888974
15	Frontal_Inf_Orb_L	0.92606791	0.92369269	0.900901654	0.911284416	0.927409539	0.916482836
16	Frontal_Inf_Orb_R	0.9237721	0.92071827	0.895152634	0.907383372	0.924997479	0.912894322
17	Rolandic_Oper_L	0.90371206	0.89862543	0.873739867	0.889072992	0.908294586	0.895813179
18	Rolandic_Oper_R	0.91948087	0.91520169	0.891323866	0.906473842	0.923096213	0.912928165
19	Supp_Motor_Area_L	0.8703661	0.86699612	0.852170361	0.857588894	0.869244057	0.862940853
20	Supp_Motor_Area_R	0.86512639	0.86016361	0.843580136	0.851272707	0.864310565	0.856253532
21	Olfactory_L	0.83978182	0.88433679	0.847352927	0.884038317	0.900095355	0.880986657
22	Olfactory_R	0.84348134	0.88587624	0.848459286	0.883339332	0.901870587	0.882599832
23	Frontal_Sup_Medial_L	0.91302599	0.90569833	0.889142748	0.894954739	0.910456821	0.896690409
24	Frontal_Sup_Medial_R	0.91384881	0.90700698	0.891467502	0.894888874	0.912325321	0.895444853

25	Frontal_Med_Orb_L	0.89918627	0.88905411	0.860802783	0.884015548	0.902084611	0.886488669
26	Frontal_Med_Orb_R	0.90261878	0.89498764	0.870195309	0.888750477	0.906404481	0.89028485
27	Rectus_L	0.91103123	0.90507338	0.869750969	0.89830088	0.917314871	0.898155032
28	Rectus_R	0.90823948	0.90197264	0.869900822	0.89507893	0.912802282	0.8949391
29	Insula_L	0.93781693	0.93613111	0.921402454	0.931086746	0.941663574	0.934415293
30	Insula_R	0.9378854	0.9354275	0.919248822	0.928650557	0.941143405	0.932911431
31	Cingulum_Ant_L	0.91440073	0.91107858	0.890874029	0.90451477	0.918306375	0.908386617
32	Cingulum_Ant_R	0.91218856	0.90955056	0.888534279	0.904449103	0.916459179	0.905896287
33	Cingulum_Mid_L	0.92125615	0.91772931	0.898202142	0.912387378	0.924317007	0.915481381
34	Cingulum_Mid_R	0.92030201	0.91727865	0.897718165	0.912289549	0.923665346	0.915510814
35	Hippocampus_L	0.92177067	0.91786398	0.894199104	0.914699889	0.927288501	0.915927373
36	Hippocampus_R	0.92248158	0.91834146	0.894813281	0.915937886	0.928417001	0.916594014
37	Amygdala_L	0.89744297	0.89426729	0.85667214	0.892066833	0.908252353	0.888928258
38	Amygdala_R	0.8909999	0.89570446	0.850670417	0.891921282	0.911259065	0.889751372
39	Calcarine_L	0.90062996	0.8942866	0.855922848	0.883147265	0.907049171	0.890006043
40	Calcarine_R	0.91235911	0.90628921	0.871711044	0.894102133	0.916461716	0.901770557
41	Cuneus_L	0.91558506	0.91029995	0.888516205	0.901249078	0.917631584	0.905583484
42	Cuneus_R	0.91763499	0.91274655	0.889807864	0.902780564	0.918135246	0.907534407
43	Lingual_L	0.91945852	0.91597651	0.889125845	0.906359002	0.92407475	0.91240192
44	Lingual_R	0.92689643	0.92378648	0.898038062	0.914619983	0.931708226	0.920930573
45	Occipital_Sup_L	0.87504093	0.86627741	0.840667771	0.855256354	0.875333213	0.859459194
46	Occipital_Sup_R	0.88792553	0.88050425	0.85169197	0.867731762	0.887820028	0.87531856
47	Occipital_Mid_L	0.89997269	0.88914415	0.8640688	0.878216076	0.899188768	0.883947328
48	Occipital_Mid_R	0.87685028	0.8674435	0.838585249	0.853816334	0.877589577	0.861707378
49	Occipital_Inf_L	0.86678897	0.85573483	0.818840896	0.840798752	0.871194957	0.851488133
50	Occipital_Inf_R	0.89659584	0.88798965	0.855995047	0.874418703	0.898878686	0.881138037
51	Fusiform_L	0.90309839	0.89924837	0.866870764	0.886244739	0.908146472	0.892796577
52	Fusiform_R	0.8966213	0.891548	0.860262156	0.878255965	0.902292866	0.887836029

53	Postcentral_L	0.91586311	0.90984498	0.891085242	0.899576946	0.913971014	0.903406451
54	Postcentral_R	0.91331578	0.90809617	0.885708965	0.894694524	0.912347065	0.900211622
55	Parietal_Sup_L	0.87716218	0.86909424	0.841455889	0.855428713	0.875198427	0.857597864
56	Parietal_Sup_R	0.87831147	0.87154855	0.841649011	0.855116197	0.87478017	0.858677128
57	Parietal_Inf_L	0.79499334	0.78253338	0.751767471	0.767257045	0.797058156	0.779489102
58	Parietal_Inf_R	0.83386438	0.82459419	0.792271938	0.805985483	0.832883427	0.81457852
59	SupraMarginal_L	0.87800527	0.86998316	0.84702318	0.859630257	0.879090623	0.868587279
60	SupraMarginal_R	0.88424169	0.87830635	0.847614199	0.863951419	0.88620091	0.874994883
61	Angular_L	0.82655018	0.81250206	0.786568666	0.802494553	0.826448215	0.813401042
62	Angular_R	0.83709775	0.8263591	0.800159013	0.815647099	0.835443635	0.823907625
63	Precuneus_L	0.92841757	0.92483407	0.905714971	0.916496257	0.93021844	0.9207075
64	Precuneus_R	0.92426747	0.92175034	0.901067747	0.912666143	0.926951687	0.917319929
65	Paracentral_Lobule_L	0.88828835	0.88466762	0.868472332	0.87823553	0.887723343	0.881767741
66	Paracentral_Lobule_R	0.89032365	0.88610598	0.867224724	0.879193105	0.89042856	0.884562368
67	Caudate_L	0.91338436	0.92667513	0.907627182	0.920060235	0.931468615	0.923010554
68	Caudate_R	0.91812157	0.92657253	0.908959243	0.919054994	0.932140306	0.923057344
69	Putamen_L	0.9390219	0.93686757	0.91983945	0.931274798	0.941924444	0.934191297
70	Putamen_R	0.93792662	0.93684045	0.917856085	0.930030069	0.942025928	0.932258886
71	Pallidum_L	0.85605022	0.91335532	0.885231205	0.910585873	0.921217057	0.906882565
72	Pallidum_R	0.8793867	0.91691361	0.892617568	0.911714999	0.922754255	0.911052848
73	Thalamus_L	0.87736006	0.95084302	0.934595231	0.944456201	0.955916	0.947931602
74	Thalamus_R	0.87106302	0.95298395	0.938297562	0.947142596	0.956823721	0.950002667
75	Heschl_L	0.90146126	0.89745082	0.86535597	0.888637335	0.905376906	0.893900398
76	Heschl_R	0.90349278	0.90004084	0.867840942	0.892698393	0.910590253	0.897159804
77	Temporal_Sup_L	0.93185619	0.9280823	0.909622402	0.918533698	0.932195551	0.924951431
78	Temporal_Sup_R	0.93081883	0.92718196	0.908013331	0.916486772	0.931839389	0.924483267
79	Temporal_Pole_Sup_L	0.91480124	0.9093237	0.883661663	0.901325863	0.916252298	0.905961273
80	Temporal_Pole_Sup_R	0.90765483	0.90397944	0.865842085	0.890458905	0.911042134	0.896409421

81	Temporal_Mid_L	0.91316275	0.90790218	0.886070057	0.896357167	0.914303635	0.904070663
82	Temporal_Mid_R	0.92282178	0.91914359	0.898466391	0.905429285	0.92377441	0.914308399
83	Temporal_Pole_Mid_L	0.89586595	0.88938552	0.852797267	0.874952936	0.898481284	0.880513276
84	Temporal_Pole_Mid_R	0.90270433	0.89673003	0.853326118	0.879358492	0.905503005	0.884764511
85	Temporal_Inf_L	0.91224974	0.90624955	0.875794501	0.889262779	0.91427741	0.897688258
86	Temporal_Inf_R	0.91601784	0.9099861	0.88166083	0.892091862	0.916706763	0.901480151
87	Cerebellum_Crus1_L	0.94309905	0.94010191	0.922953813	0.932130982	0.943436171	0.934088122
88	Cerebellum_Crus1_R	0.94845438	0.94537265	0.927285723	0.93565785	0.949424355	0.938271178
89	Cerebellum_Crus2_L	0.93291764	0.92624151	0.90048502	0.915378546	0.932810644	0.914477059
90	Cerebellum_Crus2_R	0.9243896	0.91800565	0.888445493	0.902542555	0.925035938	0.902210352
91	Cerebellum_3_L	0.88427293	0.8808562	0.845960739	0.883766912	0.897164931	0.876104977
92	Cerebellum_3_R	0.88486207	0.88781241	0.846345004	0.886952991	0.899057562	0.878977374
93	Cerebellum_4_5_L	0.92272418	0.92094621	0.898199051	0.914732954	0.928438786	0.917055001
94	Cerebellum_4_5_R	0.90456303	0.90268281	0.876649516	0.896298764	0.910320419	0.896985396
95	Cerebellum_6_L	0.93570547	0.93331679	0.912953514	0.924852335	0.938340918	0.929238127
96	Cerebellum_6_R	0.93875005	0.93707085	0.917957386	0.927855784	0.942907632	0.932160838
97	Cerebellum_7b_L	0.80606445	0.79440464	0.743508192	0.779447976	0.814044293	0.783868566
98	Cerebellum_7b_R	0.81295469	0.80152957	0.75814829	0.785233954	0.81966263	0.793588175
99	Cerebellum_8_L	0.93061659	0.92730983	0.9022169	0.918161659	0.934065921	0.919794667
100	Cerebellum_8_R	0.93457304	0.9297563	0.907124269	0.921948604	0.937017571	0.924522536
101	Cerebellum_9_L	0.93952161	0.93644237	0.915462449	0.929800829	0.943877646	0.932563286
102	Cerebellum_9_R	0.93540429	0.931885	0.908322037	0.923865178	0.938789156	0.927798001
103	Cerebellum_10_L	0.88731824	0.87745959	0.847481976	0.880515617	0.894370957	0.869658955
104	Cerebellum_10_R	0.89738617	0.8918664	0.862372569	0.888251728	0.901960234	0.88553432
105	Vermis_1_2	0.85286368	0.84944366	0.808505328	0.865250343	0.873591993	0.841630677
106	Vermis_3	0.88031453	0.89339164	0.86015436	0.900287587	0.91442588	0.893047241
107	Vermis_4_5	0.92301205	0.9251845	0.904576866	0.921531539	0.933487694	0.92164322

108	Vermis_6	0.8903081	0.88862238	0.854158034	0.883589695	0.900148275	0.881459079
109	Vermis_7	0.88424936	0.88430543	0.846991555	0.879644931	0.894511129	0.876049088
110	Vermis_8	0.91528095	0.91144512	0.889030895	0.909490428	0.92125012	0.907658172
111	Vermis_9	0.89885087	0.89376478	0.85673071	0.895248569	0.909065079	0.890177684
112	Vermis_10	0.88582519	0.88176634	0.85010123	0.888230494	0.893759661	0.875679534
113	Pons	0.96646647	0.96554829	0.958595393	0.962104001	0.968296002	0.964560776
114	Dorsal_Mesopontine	0.82149933	0.90879691	0.891646798	0.908348024	0.91744528	0.90444335
115	Entorhinal_Cortex_L	0.89849056	0.89423825	0.863935997	0.888468253	0.904972508	0.889677729
116	Entorhinal_Cortex_R	0.88792097	0.88367833	0.84633356	0.878500773	0.896153938	0.878224579
117	ParaHippocampal_L	0.91144567	0.90790963	0.880309664	0.904057232	0.915458507	0.904501211
118	ParaHippocampal_R	0.90972812	0.90679542	0.871664165	0.900191842	0.917015277	0.902004042
119	Cingulum_Post_L	0.89157975	0.89007629	0.853863854	0.88161995	0.898935928	0.885530139
120	Cingulum_Post_R	0.89885709	0.89634997	0.862707041	0.891346704	0.905767533	0.894675214
121	Retrosplenial_Cortex_L	0.89665472	0.91158482	0.87959647	0.906484583	0.922356673	0.908102736
122	Retrosplenial_Cortex_R	0.87159952	0.90038525	0.861322625	0.895242477	0.910932436	0.900259311

Supplemental Table 2: The G5th Percentile Hausdorff Distance of the Test Dataset

Classes#	Name of Substructures	nnU-Net V2	SegResNet	UNETR	SwinUNETR	U-Mamba_Bot	U-Mamba_Enc
1	Precentral_L	2.085633829	2.162502212	3.380359789	2.679966664	4.172767325	2.626088573
2	Precentral_R	2.053815334	2.090504992	2.486612236	2.326139426	4.102392081	2.201455215
3	Frontal_Sup_L	2.497753852	2.574450497	3.406415869	2.954133924	5.30498111	2.744098101
4	Frontal_Sup_R	2.553886054	2.663241231	3.011411961	2.901574415	5.12578998	2.842200298
5	Frontal_Sup_Orb_L	1.373408658	1.452890258	2.146156695	1.624863059	2.703766979	1.511002359
6	Frontal_Sup_Orb_R	1.465054352	1.578384036	2.069822246	1.693284063	2.868428898	1.557612404
7	Frontal_Mid_L	2.732795123	2.827560996	3.987127699	3.414023415	5.911726558	3.071336263

8	Frontal_Mid_R	2.71584561	2.834972431	3.205910342	3.195379844	5.538656988	3.0405731
9	Frontal_Mid_Orb_L	2.01175051	2.096160151	3.222393232	2.609782381	4.217298033	2.310302848
10	Frontal_Mid_Orb_R	1.706930546	1.791112856	2.321699695	2.024562097	3.37126712	1.790640471
11	Frontal_Inf_Oper_L	1.959601413	2.032198534	3.509557679	2.700067651	4.019062758	2.368434194
12	Frontal_Inf_Oper_R	2.019417323	2.156641399	2.541266689	2.404332617	3.946320524	2.074601781
13	Frontal_Inf_Tri_L	2.448871981	2.391536976	3.970229203	3.098186599	4.899361128	2.715858204
14	Frontal_Inf_Tri_R	2.299886803	2.428234411	2.71941565	2.744131793	4.601421251	2.39459214
15	Frontal_Inf_Orb_L	1.561311187	1.586665664	3.041040248	2.215500903	3.075088021	1.887117261
16	Frontal_Inf_Orb_R	1.567358505	1.61248989	2.139908236	1.917661892	3.105333227	1.644880603
17	Rolandic_Oper_L	1.389219779	1.446525277	3.14274743	2.146437323	2.694214211	1.94072427
18	Rolandic_Oper_R	1.420864413	1.485814061	2.037094189	1.77143545	2.739969419	1.649388718
19	Supp_Motor_Area_L	3.391568143	3.462768146	3.774820683	3.620539348	6.899234712	3.453129888
20	Supp_Motor_Area_R	3.509318046	3.612086501	3.817653433	3.685928557	7.112391538	3.611503479
21	Olfactory_L	1.325302815	1.071365695	1.371153207	1.158875566	2.077839167	1.108270486
22	Olfactory_R	1.26564579	1.059152051	1.291634525	1.100749874	2.097267589	1.100785719
23	Frontal_Sup_Medial_L	2.701568098	2.80525986	3.058340716	2.909440906	5.397516244	2.782145538
24	Frontal_Sup_Medial_R	2.795098521	2.911137782	3.042735076	2.978415808	5.544865756	2.912547246
25	Frontal_Med_Orb_L	1.524163344	1.693423999	2.112007995	1.751107788	3.008557075	1.65107908
26	Frontal_Med_Orb_R	1.555414911	1.685803884	1.964810825	1.707720903	3.061161777	1.667355829
27	Rectus_L	1.205296804	1.243979599	1.495155082	1.288667103	2.306241454	1.276082924
28	Rectus_R	1.170084705	1.231171047	1.473857996	1.284443643	2.295862702	1.302538567
29	Insula_L	1.173818235	1.179290563	2.090677235	1.786746869	2.23743004	1.433883036
30	Insula_R	1.141807972	1.174230492	1.393672461	1.355920375	2.19610069	1.239689225
31	Cingulum_Ant_L	1.278256468	1.327096246	1.579981252	1.445474077	2.476970791	1.345297072
32	Cingulum_Ant_R	1.386797643	1.429987196	1.605222393	1.460677704	2.715993834	1.456752111
33	Cingulum_Mid_L	1.182068282	1.221415346	1.402300631	1.299338604	2.326263514	1.227058931
34	Cingulum_Mid_R	1.26072089	1.287761484	1.46764591	1.352133966	2.455919207	1.319403078
35	Hippocampus_L	1.035228401	1.042434121	1.903032064	1.360114978	2.050281974	1.216733713

36	Hippocampus_R	1.035707067	1.04059894	1.354330598	1.146829209	2.043098614	1.142058336
37	Amygdala_L	1.072843967	1.076991411	1.894616932	1.359890415	2.089549803	1.271729595
38	Amygdala_R	1.146386288	1.099583268	1.517323363	1.195061673	2.080488377	1.116195107
39	Calcarine_L	1.497181279	1.583313173	2.235348376	1.822093641	2.845801194	1.626337542
40	Calcarine_R	1.369684412	1.443289323	1.864897538	1.594364141	2.654632205	1.480921462
41	Cuneus_L	1.715116209	1.782256481	2.236218982	1.973279804	3.368684569	1.846288234
42	Cuneus_R	1.672569866	1.750710334	2.113999014	1.915734564	3.333788516	1.797800815
43	Lingual_L	1.435069007	1.486098869	2.119036793	1.742776566	2.711072307	1.597895589
44	Lingual_R	1.394200172	1.451516298	1.89245447	1.620683748	2.89855725	1.479197571
45	Occipital_Sup_L	2.129071505	2.256015191	2.94617791	2.580007225	4.35828897	2.340174205
46	Occipital_Sup_R	2.067274377	2.150641665	2.702241995	2.39596587	4.099022867	2.246825091
47	Occipital_Mid_L	2.399151027	2.596521609	3.734548674	3.044311266	4.948373692	2.795275618
48	Occipital_Mid_R	2.571770466	2.762444868	3.385521663	3.007806533	5.029650617	2.781391283
49	Occipital_Inf_L	2.020318218	2.145314637	3.274228854	2.660559657	3.898449568	2.383594869
50	Occipital_Inf_R	1.875899515	2.05896085	2.622245208	2.228345405	3.62732626	2.071939736
51	Fusiform_L	1.42277186	1.462178623	2.731780507	2.030007372	2.724609543	1.895161858
52	Fusiform_R	1.424572719	1.481928048	2.096928574	1.75147522	2.719048869	1.617573249
53	Postcentral_L	2.15778016	2.295208188	3.469242115	2.700997683	4.584464528	2.741614976
54	Postcentral_R	2.103334219	2.182005611	2.732738274	2.406730941	4.182247165	2.272953919
55	Parietal_Sup_L	3.109882797	3.217270391	4.042592879	3.548720926	6.341739735	3.431887571
56	Parietal_Sup_R	3.013045921	3.116734766	3.550981163	3.334901065	6.016295995	3.1883677
57	Parietal_Inf_L	3.079322339	3.232121445	4.64053766	3.72652892	6.400931768	3.693565537
58	Parietal_Inf_R	3.067380042	3.16111964	3.840187649	3.471228109	6.046698177	3.277031655
59	SupraMarginal_L	2.608713754	2.719444889	4.240346035	3.312404643	5.467104524	3.20388932
60	SupraMarginal_R	2.40098513	2.498864076	3.190882246	2.82902125	4.723494132	2.666324353
61	Angular_L	3.266791415	3.4134196	4.62285496	3.941786604	6.786738673	3.862188834
62	Angular_R	3.190947594	3.41047981	4.049390443	3.559360628	6.397739696	3.400778791
63	Precuneus_L	1.730540851	1.781651223	2.15788669	1.938382413	3.38893852	1.831296881

64	Precuneus_R	1.800551553	1.849447703	2.18161761	1.974228273	3.485216292	1.888959468
65	Paracentral_Lobule_L	2.463630165	2.556721211	2.801439852	2.671432694	5.088228118	2.559546677
66	Paracentral_Lobule_R	2.482018229	2.571977429	2.879166458	2.656391819	5.117920147	2.610568299
67	Caudate_L	1.16158176	1.105653569	1.335780818	1.380318651	2.139586763	1.18776244
68	Caudate_R	1.153232016	1.088288303	1.231435531	1.229643511	2.152036668	1.18396217
69	Putamen_L	1.104764426	1.110830876	1.642345528	1.468440778	2.127774475	1.231695781
70	Putamen_R	1.091241577	1.095098553	1.332854214	1.272808804	2.104191859	1.181644096
71	Pallidum_L	1.232535846	1.025668443	1.309955818	1.2201822	2.021826598	1.114431387
72	Pallidum_R	1.256520438	1.026631201	1.190602582	1.118491083	2.023522118	1.080600316
73	Thalamus_L	2.090868891	1.04866224	1.410412756	1.213682364	2.171809396	1.116826741
74	Thalamus_R	2.284719413	1.060954972	1.251393484	1.145608598	2.083203373	1.131515222
75	Heschl_L	1.14763253	1.160645169	2.328261755	1.864929444	2.25327163	1.497209352
76	Heschl_R	1.130037869	1.149805243	1.626825088	1.353357908	2.173795125	1.318677171
77	Temporal_Sup_L	1.653371733	1.729402612	3.650710408	2.576899644	3.658482229	2.44268448
78	Temporal_Sup_R	1.571955684	1.632419312	2.419789339	2.010684644	3.114954902	1.862924791
79	Temporal_Pole_Sup_L	1.228087125	1.274977366	2.516589872	1.818192047	2.434317082	1.560527832
80	Temporal_Pole_Sup_R	1.283794372	1.38035949	1.957221713	1.627036484	2.49460802	1.528659097
81	Temporal_Mid_L	1.940082062	2.043622575	4.401015941	2.955433307	4.175378965	2.753820964
82	Temporal_Mid_R	1.89616956	1.977126827	2.80648075	2.424742698	3.729448488	2.198090023
83	Temporal_Pole_Mid_L	1.519046318	1.581611212	2.957290166	1.951986812	2.926779243	1.937473942
84	Temporal_Pole_Mid_R	1.474522371	1.521651042	2.353270055	1.90049577	2.833464754	1.731890999
85	Temporal_Inf_L	1.60582991	1.67046563	3.840677373	2.558415763	3.447692337	2.346192432
86	Temporal_Inf_R	1.640316421	1.86616892	2.503327178	2.126181515	3.174331874	1.957809447
87	Cerebellum_Crus1_L	1.190493347	1.212679992	1.794024258	1.579430784	2.363016811	1.469406847
88	Cerebellum_Crus1_R	1.173554072	1.199762408	1.540410335	1.377469869	2.587568462	1.343177148
89	Cerebellum_Crus2_L	1.249643255	1.462217062	1.96452937	1.584491333	2.523472734	1.553870856
90	Cerebellum_Crus2_R	1.257934145	1.34400266	1.697731604	1.530438893	2.458375615	1.492203654

91	Cerebellum_3_L	1.029109711	1.034502867	1.312202386	1.05842971	2.020336299	1.064866483
92	Cerebellum_3_R	1.037466373	1.048051354	1.214387864	1.066679348	2.016282567	1.072625886
93	Cerebellum_4_5_L	1.175294322	1.187585062	1.690508359	1.303630634	2.243714512	1.255191197
94	Cerebellum_4_5_R	1.198312386	1.265786522	1.432767325	1.297551643	2.294292567	1.281252154
95	Cerebellum_6_L	1.195444497	1.355589449	1.750672218	1.524506129	2.353637809	1.338427906
96	Cerebellum_6_R	1.187672113	1.327850349	1.510517743	1.365262342	2.261911129	1.312097182
97	Cerebellum_7b_L	1.454065186	1.557003779	2.247321054	1.849421749	2.812707457	1.737630967
98	Cerebellum_7b_R	1.462088256	1.540416217	1.91718671	1.702674928	2.806549276	1.615480888
99	Cerebellum_8_L	1.309647048	1.348589319	1.986049816	1.63031351	2.486570544	1.53421137
100	Cerebellum_8_R	1.330383714	1.42389875	1.796969444	1.543612699	2.556257368	1.477297793
101	Cerebellum_9_L	1.151589405	1.187949108	1.650060634	1.288566935	2.196286384	1.24117012
102	Cerebellum_9_R	1.151740065	1.179978344	1.427470939	1.29029067	2.21500255	1.213253842
103	Cerebellum_10_L	1.030509922	1.06674654	1.483561036	1.180473422	2.038192066	1.132037718
104	Cerebellum_10_R	1.022414216	1.028538183	1.232865528	1.071722528	2.036850132	1.102670436
105	Vermis_1_2	1.033781563	1.032444009	1.097170996	1.032679452	2.037363022	1.047331249
106	Vermis_3	1.141432179	1.086337213	1.223096075	1.080164645	2.082764884	1.089612569
107	Vermis_4_5	1.226027425	1.212700732	1.259594452	1.225768801	2.369313495	1.226930961
108	Vermis_6	1.218680605	1.225127243	1.340073366	1.2467121	2.352030685	1.231295015
109	Vermis_7	1.215912842	1.338106988	1.324857048	1.239441949	2.287713453	1.243666481
110	Vermis_8	1.139939222	1.169187801	1.240135081	1.165207342	2.209087882	1.177230497
111	Vermis_9	1.059479891	1.085162996	1.263848064	1.083020859	2.069330737	1.10303607
112	Vermis_10	1.013151576	1.013904606	1.050704032	1.012228812	2.019429405	1.021737869
113	Pons	1.014452285	1.011616326	1.199815305	1.050066681	2.019737693	1.0154516
114	Dorsal_Mesopontine	1.354411738	1.014700273	1.041285355	1.021914925	2.015519154	1.023309759
115	Entorhinal_Cortex_L	1.1683272	1.191654791	2.074014456	1.378211945	2.246281278	1.367197145
116	Entorhinal_Cortex_R	1.185300583	1.21734551	1.593497597	1.316087595	2.258695014	1.313842545
117	ParaHippocampal_L	1.065931281	1.084427548	1.614544977	1.313497662	2.093679142	1.185478499
118	ParaHippocampal_R	1.06374801	1.149708534	1.457105435	1.212054655	2.093384161	1.197438374

119	Cingulum_Post_L	1.14217343	1.144074788	1.424574828	1.235490249	2.228011544	1.160788933
120	Cingulum_Post_R	1.166391674	1.176771482	1.458877042	1.225902534	2.272397962	1.20015226
121	Retrosplenial_Cortex_L	1.191836934	1.096151671	1.338345906	1.143343258	2.107284138	1.129845527
122	Retrosplenial_Cortex_R	1.233701353	1.072938299	1.296829944	1.10889766	1.047103505	1.104662046

Supplemental Table 3: TRIPODAI checklist

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title page
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Page 1, Abstract
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Page 2-3
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	Page 3 Not Applicable
	3c	D;E	Describe any known health inequalities between sociodemographic groups	Page 4
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	
METHODS				Page 4
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	Page 5
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Page 5 and Table 1
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	Page 5 Not Applicable
	6b	D;E	Describe the eligibility criteria for study participants	
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	Page 5
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across	Not Applicable

Not Applicable

Not Applicable

Page 5-6

			relevant sociodemographic groups	
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	**Page 7**
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	**Not Applicable**
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	**Page 5**
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	**Page 5**
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	**Page 5**
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	**Page 6**
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	**Page 6**
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations[3]	**Not Applicable**
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	**Page 6-7**
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	**Page 8**
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	**Page 7**
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	**Not Applicable**
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	**Not Applicable**

Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Page 8

[1] D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

[2] Separately for all model building approaches.

[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	Page 21, Table 1
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	Not Applicable
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	Not Applicable
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	Page 2
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Page 15
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	Not Applicable
Data sharing	18e	D;E	Provide details of the availability of the study data	Page 5,15
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	Page 6
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	No involvement
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Not Applicable
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	Page 21, Table 1

	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	Page 10
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Page 9, 10
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	Page 8
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	Page 9, 10
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	Not examined
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Page 10
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	Page 13
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	Page 14
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	Page 14
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	Page 13, 14

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378