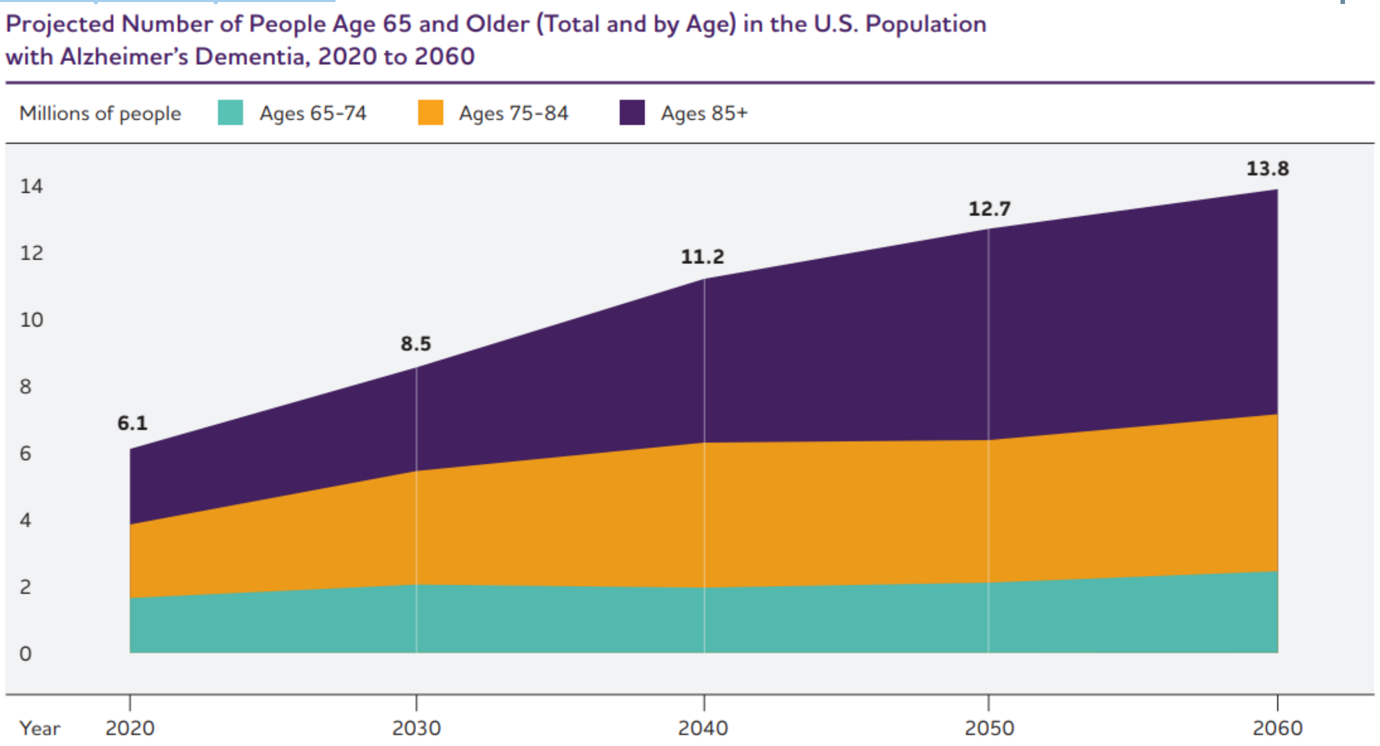
OAS-52:
Alzheimer's Disease Detection Using MRI and the ResNet-50 Architecture

BMEN 619 Final Project

Brooke Kindleman

Background[1]

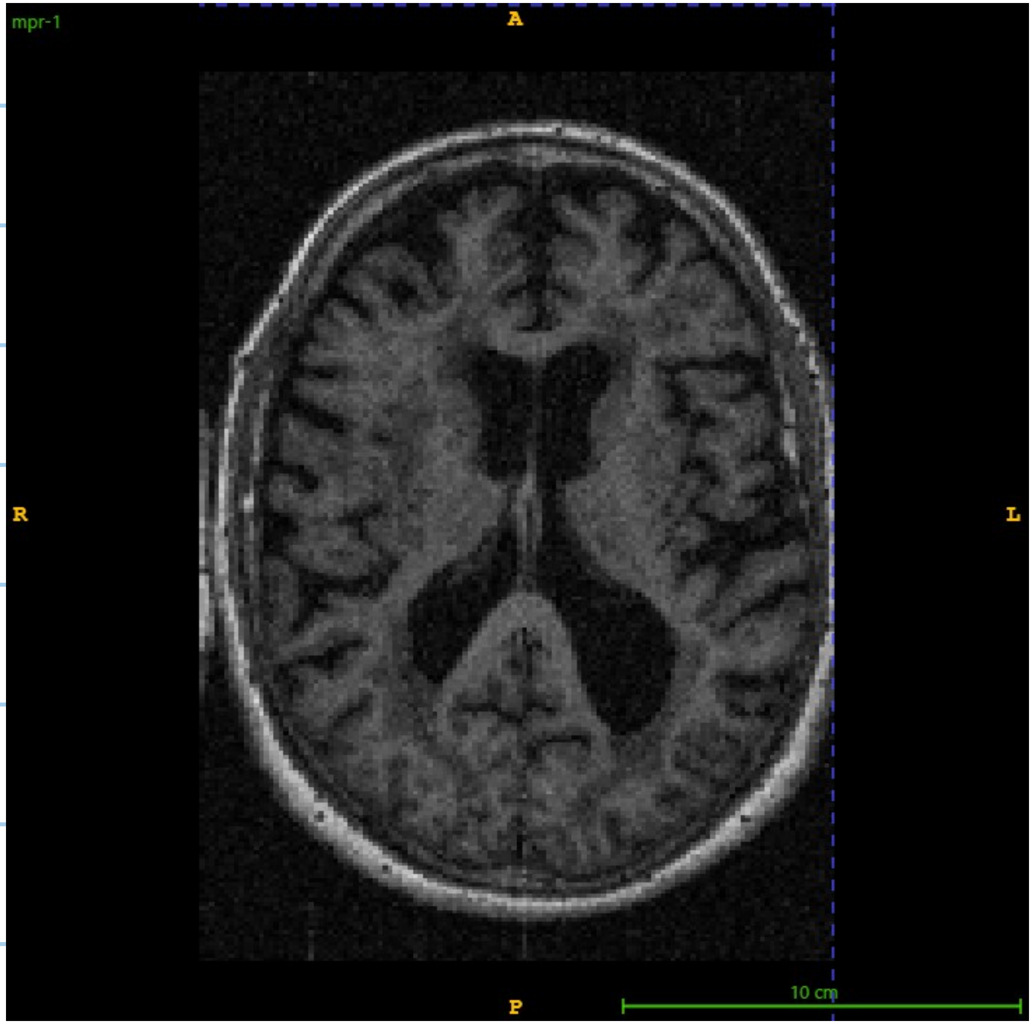
- Alzheimer's disease (AD) is an incurable neurodegenerative condition that leads to cognitive deficits, memory loss, severe disability, and can lead to death
- The lifetime risk for developing AD is approximately **1 in 5 (20%) for women** and **1 in 10 (10% in men),** after the age of 45
- Early detection of AD can give researchers the opportunity to investigate early-intervention treatments that could slow or halt disease progression



Source: [1]

Background Cont.

- AD results in structural changes to the brain including [2]:
  - Reduced grey matter volume
  - Cerebral atrophy
- Structural changes in the brain can be detected using T1-weighted magnetic resonance imaging (MRI) [3]
- Machine learning models can be used to recognize structural changes in MRI acquisitions [3]



T1-Weighted MRI of a male patient aged 79 with AD. Source: [4-5]

Primary Goal

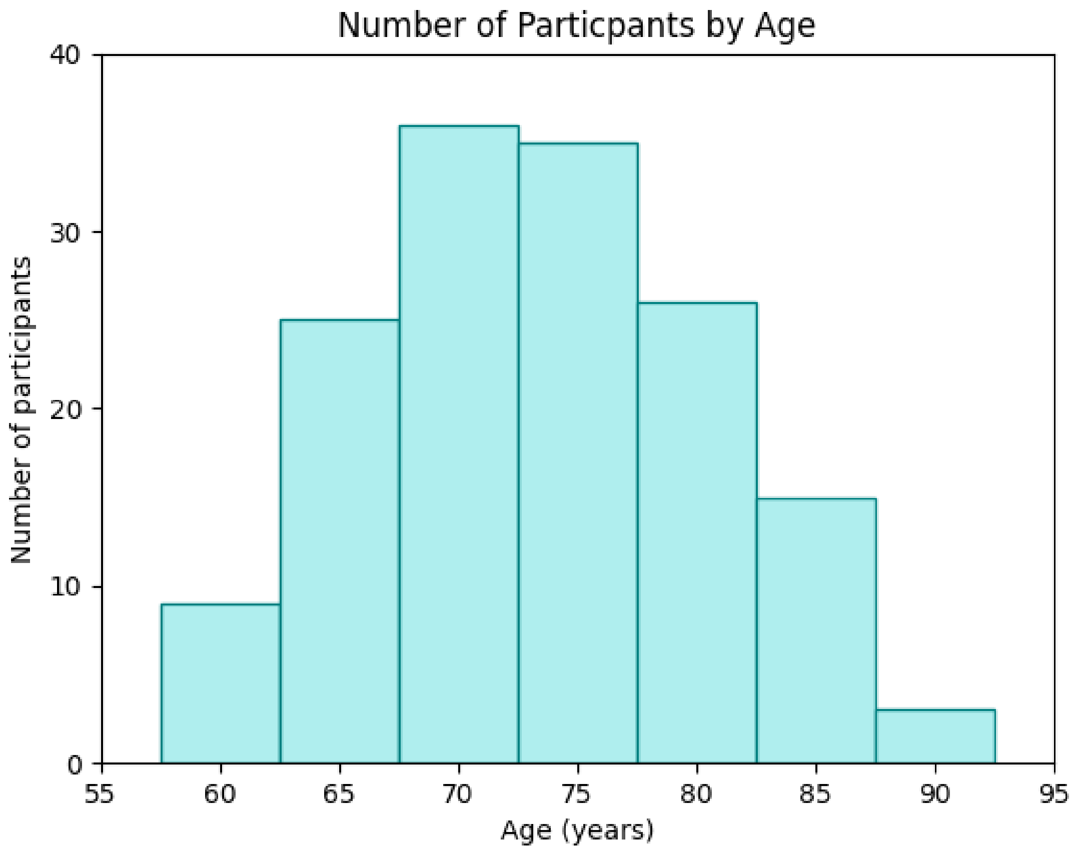
- Train and evaluate a binary classification model capable of diagnosing AD using only MRI acquisitions

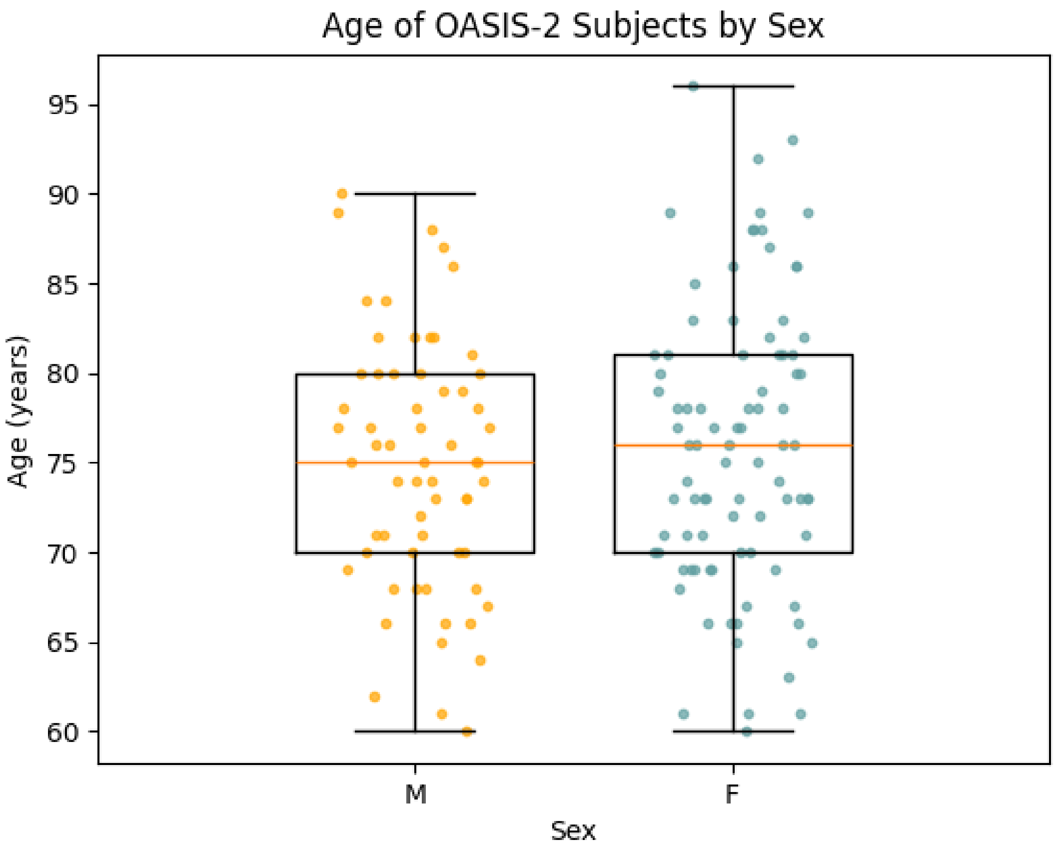
Methods

OASIS-2 Dataset

- Open Access Series of Imaging Studies, subset 2 **(OASIS-2)**[5]
  - **Longitudinal** dataset containing **MRIs** of subjects **with and without AD**
- Contains **150 subjects**, from between age **60-96**
- Subjects were imaged at between **2-5 visits**, with **at least 2** acquisitions collected at each visit
- Dataset contains a total of **1873 MRI acquisitions**
- OASIS-2 is available for **research purposes**, with explicit limitations on use in facial reconstruction

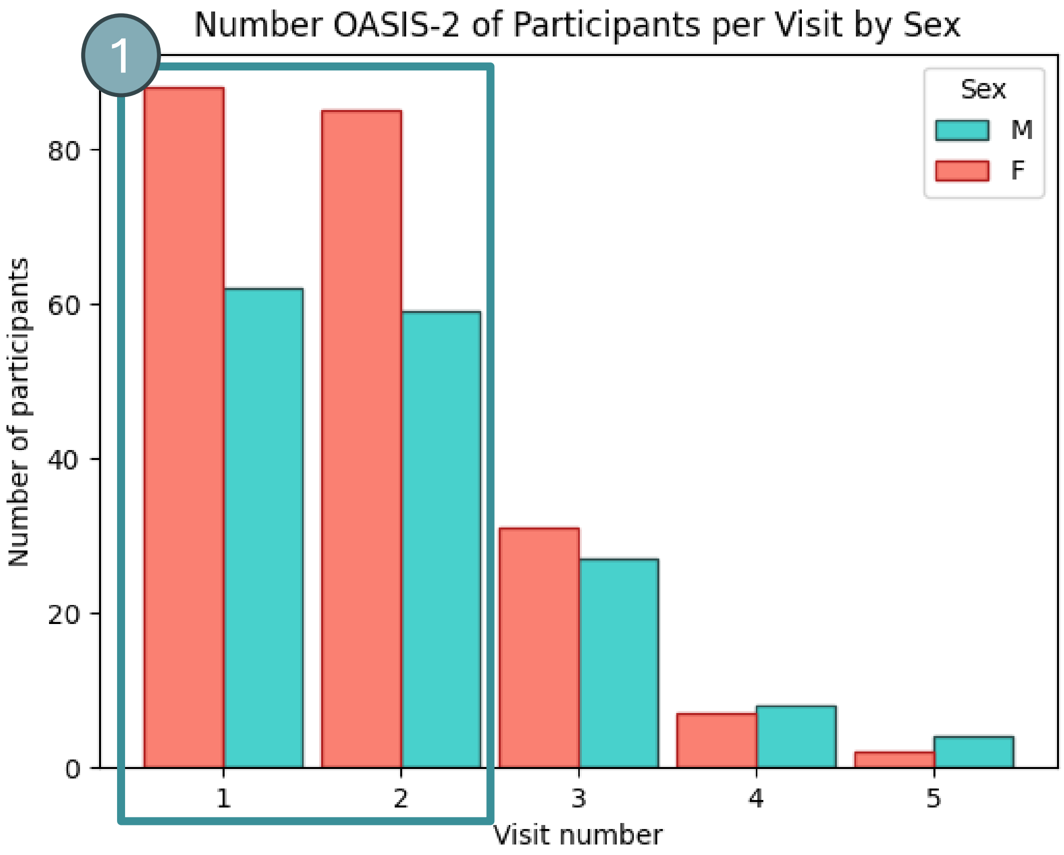
Demographics

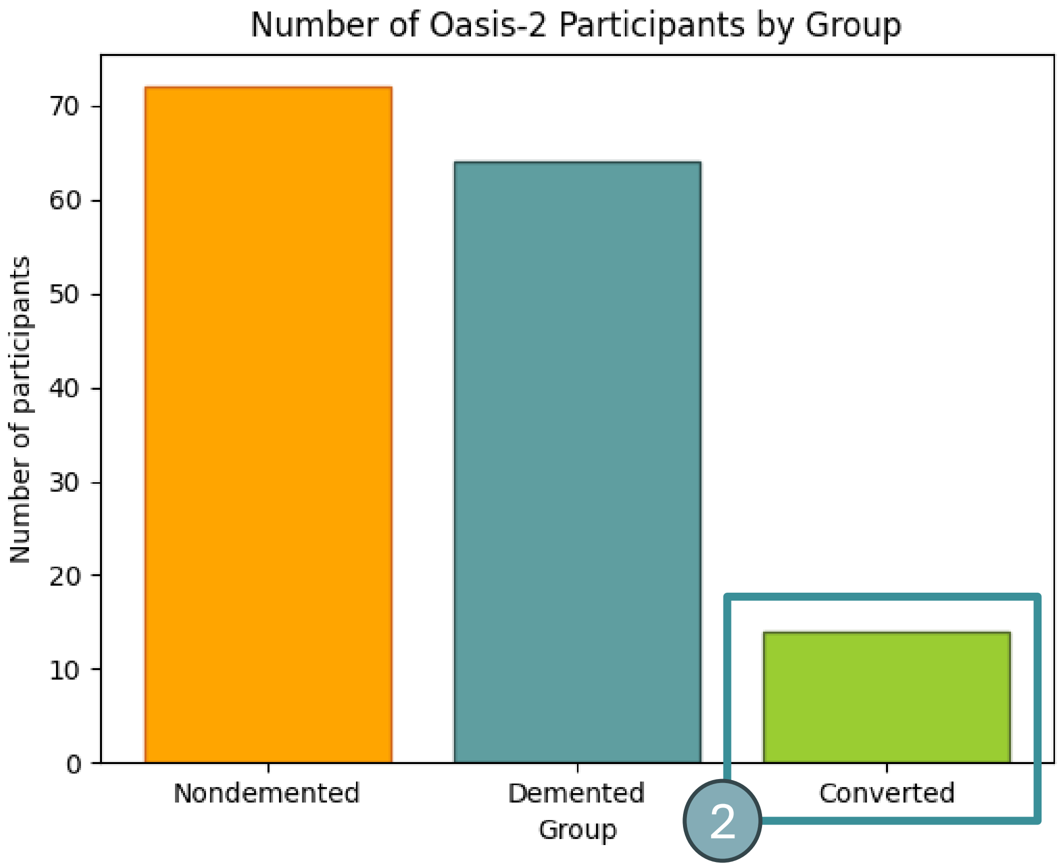




- Age is normally distributed
- No significant difference in mean age between male and female subjects

Demographics





Primary Issues with OASIS-2

1. Lack of Participants at Visits 3-5

- **Less than 50%** of participants attended visits 3-5
- Male-to-female **ratio changes** from visits 1-2 to 3-5

Remove all acquisitions from visits 3-5

2. Converted Class

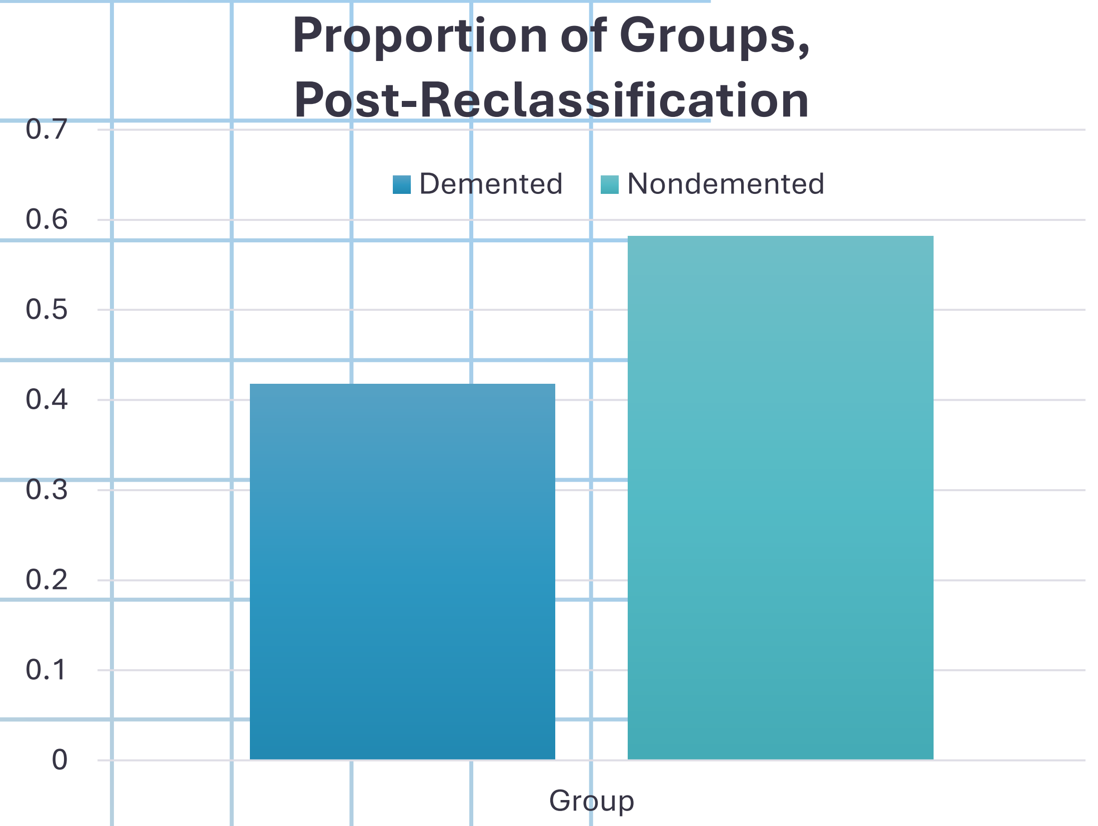
- Contains participants who began the study **without** AD, but **developed it** at some point during before the study finished
- Classified as converted for **all imaging sessions**, regardless of condition at the time of the scan

Removing the Converted Class

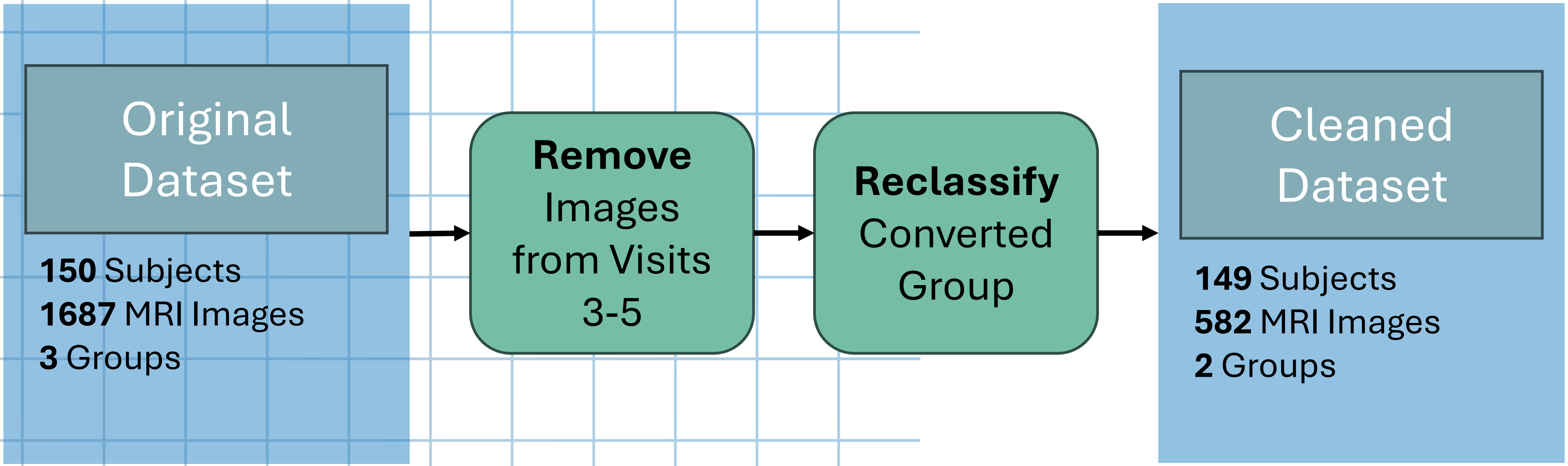
- There is a **strong correlation** (ρ=0.887) between Clinical Dementia Rating (**CDR**) and non-converted groups

$$\text{Binary Classification} = \begin{cases} \text{Demented} & \text{if CDR} \geq 1.5 \\ \text{Nondemented} & \text{if CDR} < 1.5 \end{cases}$$

- **Solution:** Reclassify subjects in the converted group per-visit, based on CDR

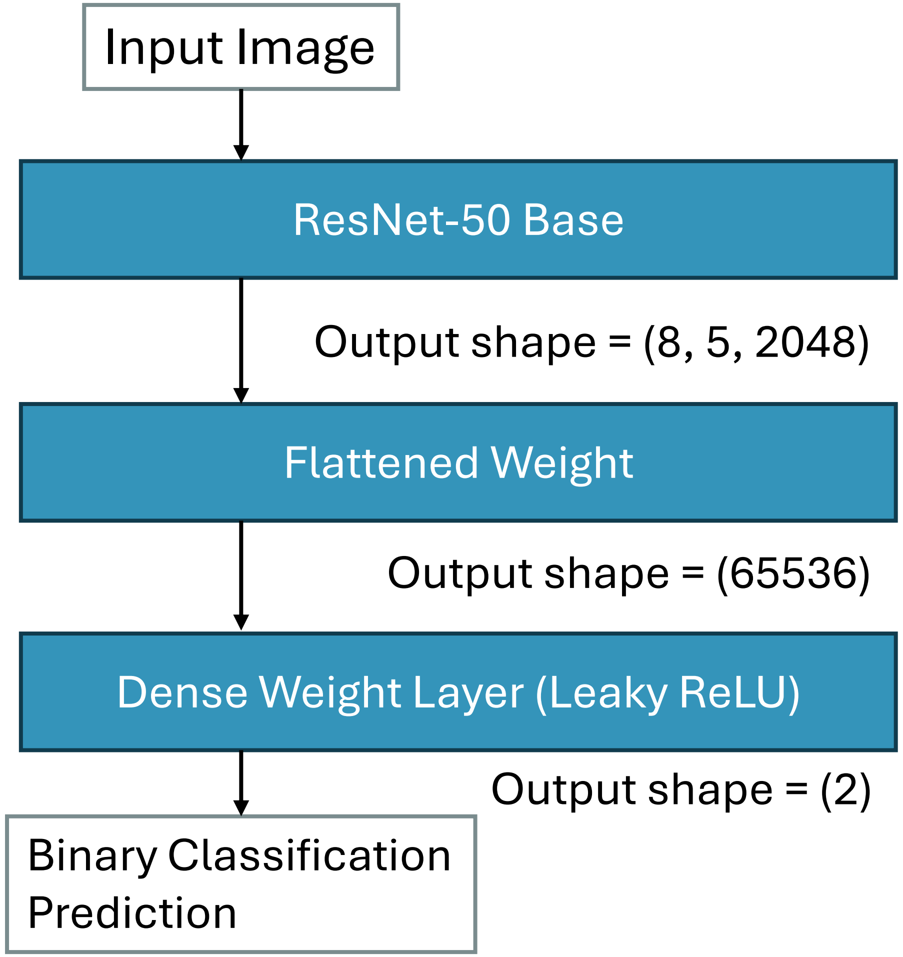


Data Cleaning Procedure



OAS-52 Architecture

- Utilizes ResNet-50[6], a **residual neural network (RNN)** developed for image recognition
- OAS-52 adds **two** additional layers to the ResNet-50 base
  - Shape flattening reduces dimensionality
  - Dense layer used to provide a two-class prediction output
    - Uses Leaky Rectified Linear Unit (ReLU)



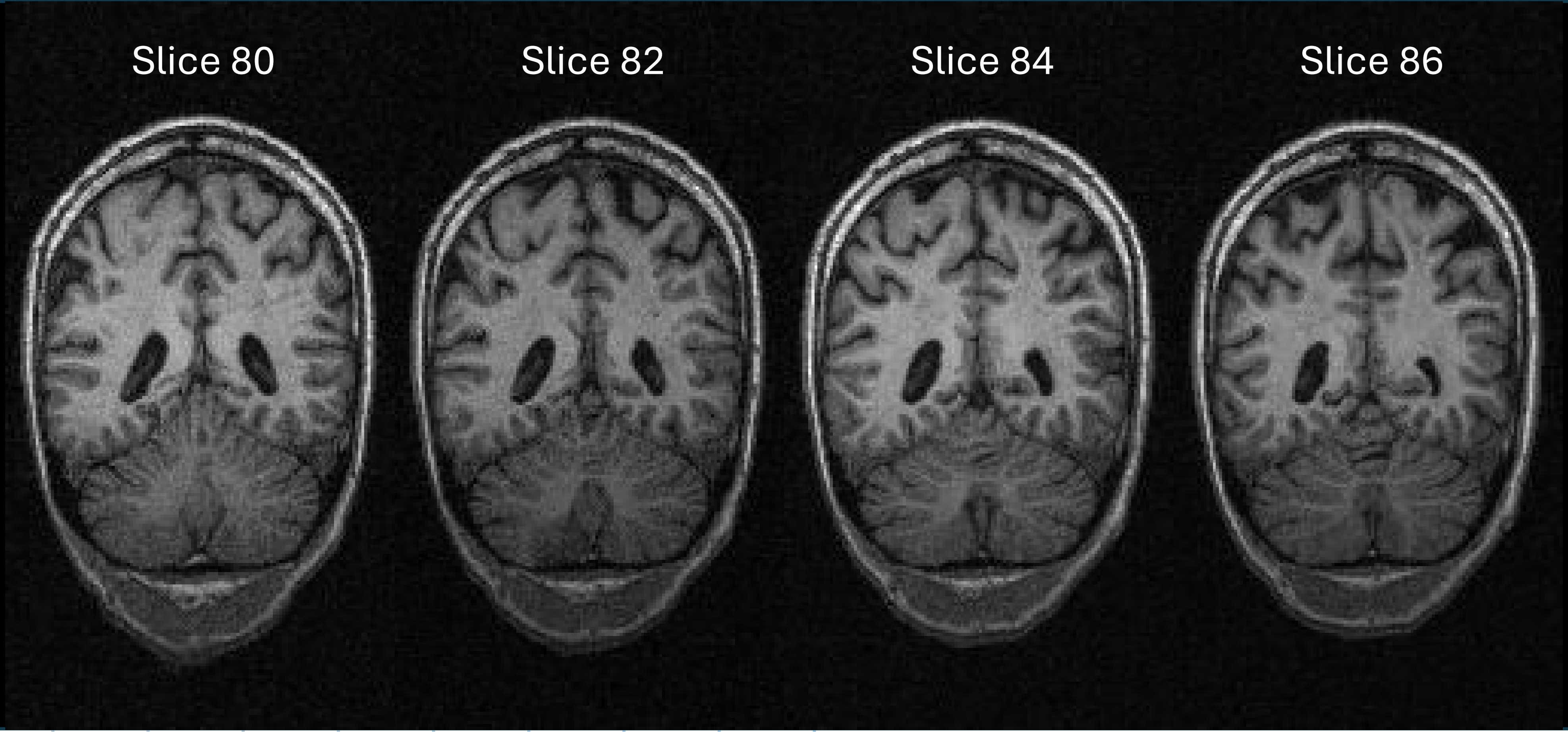
Preprocessing

- OASIS-2 contains .nifiti**.img 3-D images**
  - Images come pre-registered to the same atlas
- Resnet-50 architecture requires **2-D images**

1. Convert each .nifti.img acquisition into **4 coronal slices** (slices = {80, 82, 84, 86)
2. Normalize each image using **MinMax**

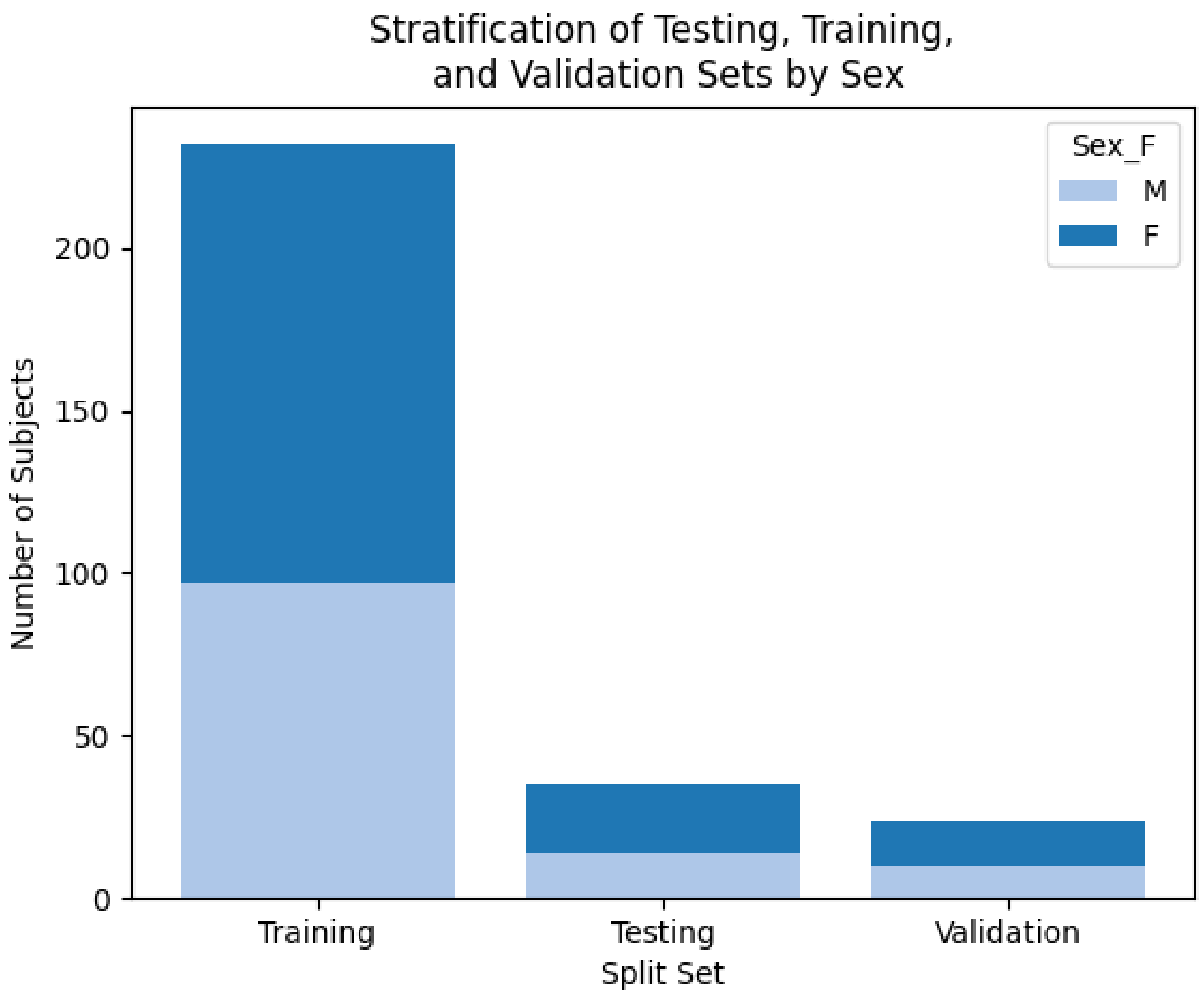
582 3-D MRI Acquisitions → **2,328** 2-D Slices

Slices Produced Per-Subject



TVT-Split Sets

Set	Number of Images (%)
Training	1856 (80%)
Testing	280 (12%)
Validation	192 (8%)



Training and Hyperparameters

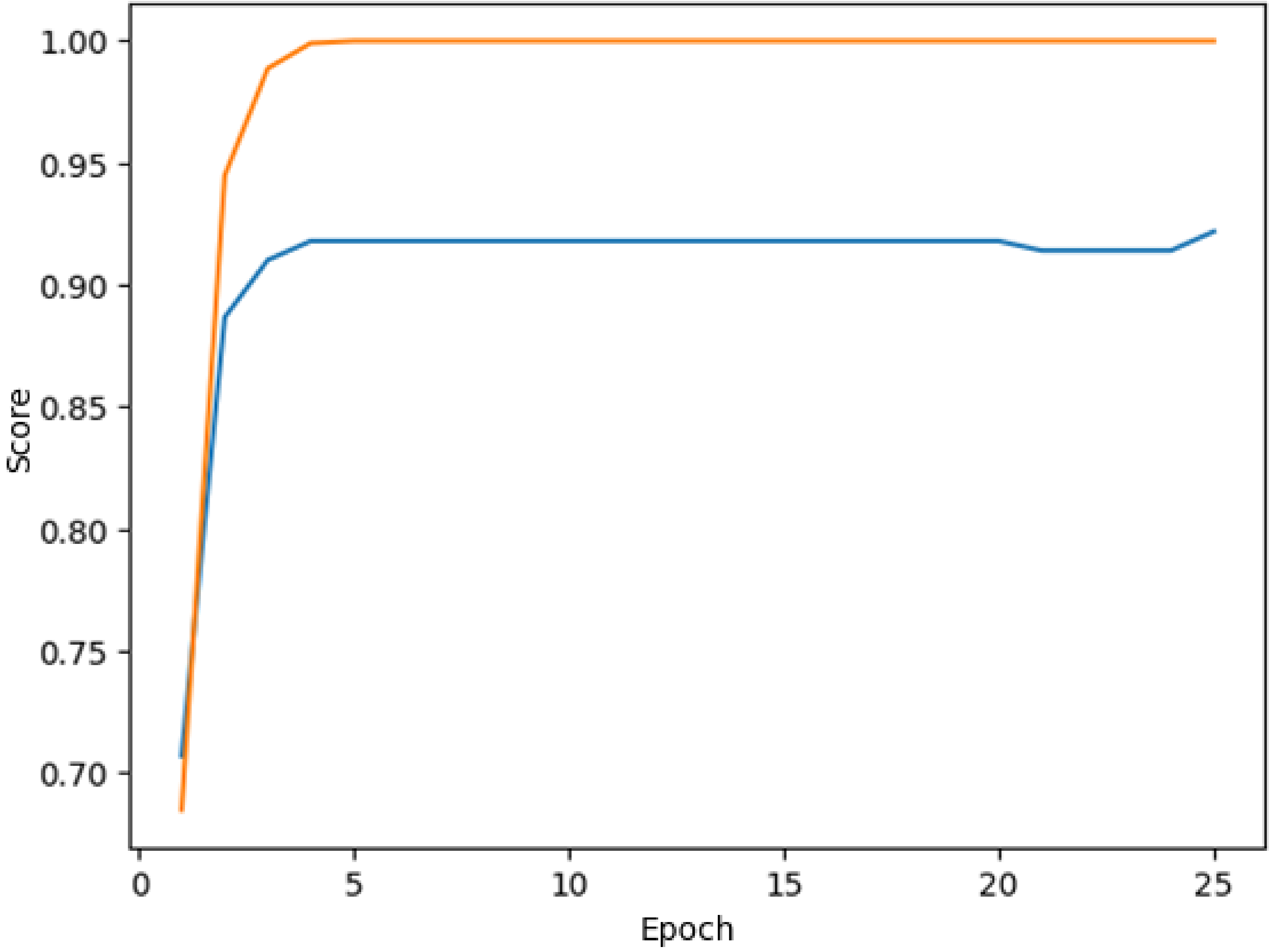
Hyperparameters

Hyperparameter	Value
Optimizer	Adam
Loss	Sparse categorical crossentropy
Input data shape	(240, 128, 3)
Epochs	25

Hardware

Hardware	Value
GPU	Nvidia GeForce RTX 3080
VRAM	7810 MB

Training and Validation Accuracy



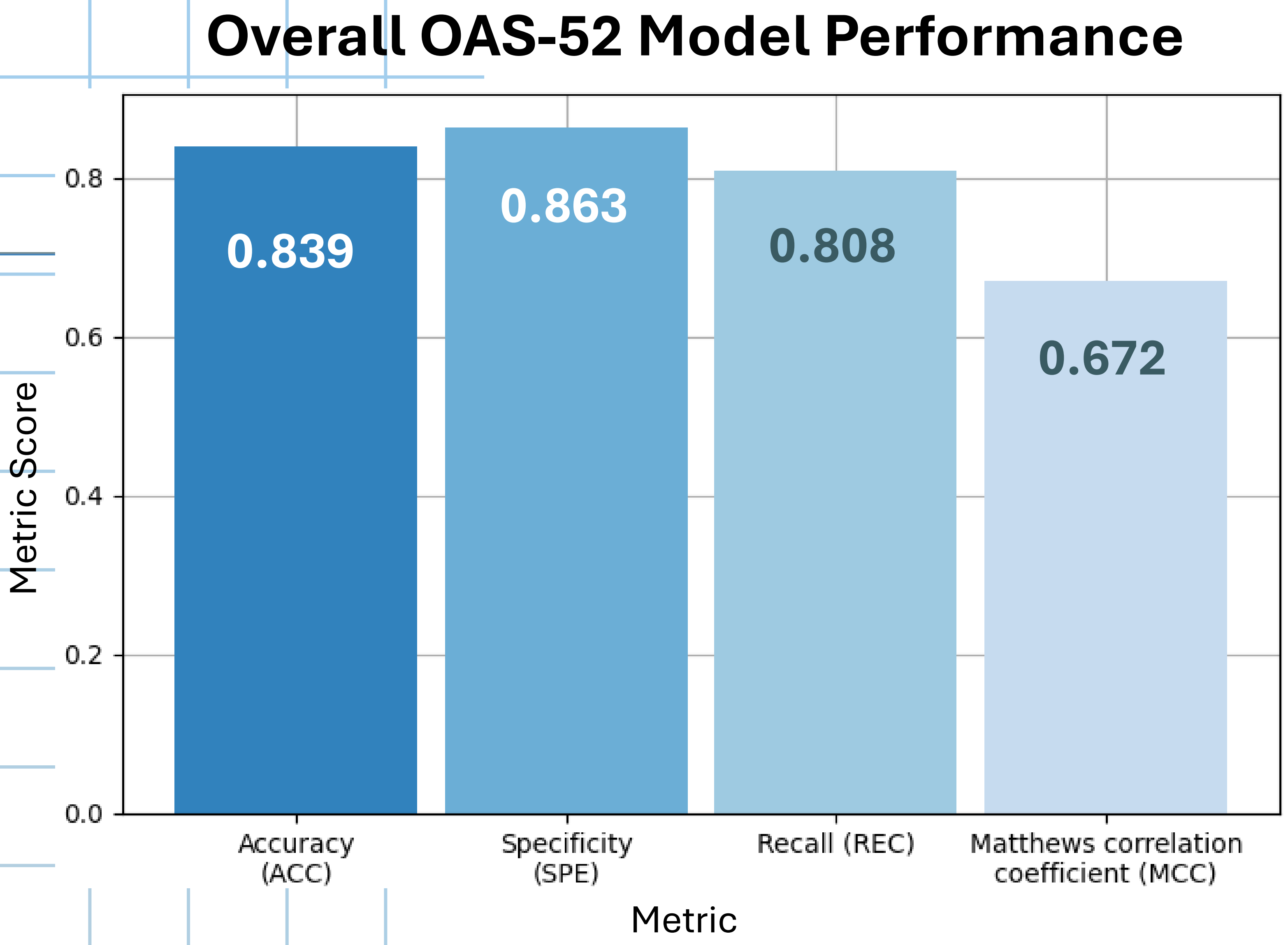
Evaluation Metrics

- **3 established** metrics for machine learning evaluation
  - Accuracy (ACC)
  - Specificity (SPE)
  - Recall (REC)
- Matthews correlation coefficient (MCC) [7]
  - Specific for use in evaluating correlations between binary classes
- **4 fairness** metrics [-9]
  - Equal odds difference (EOD)
  - Average odds difference (AOD)
  - Statistical parity difference (SPD)

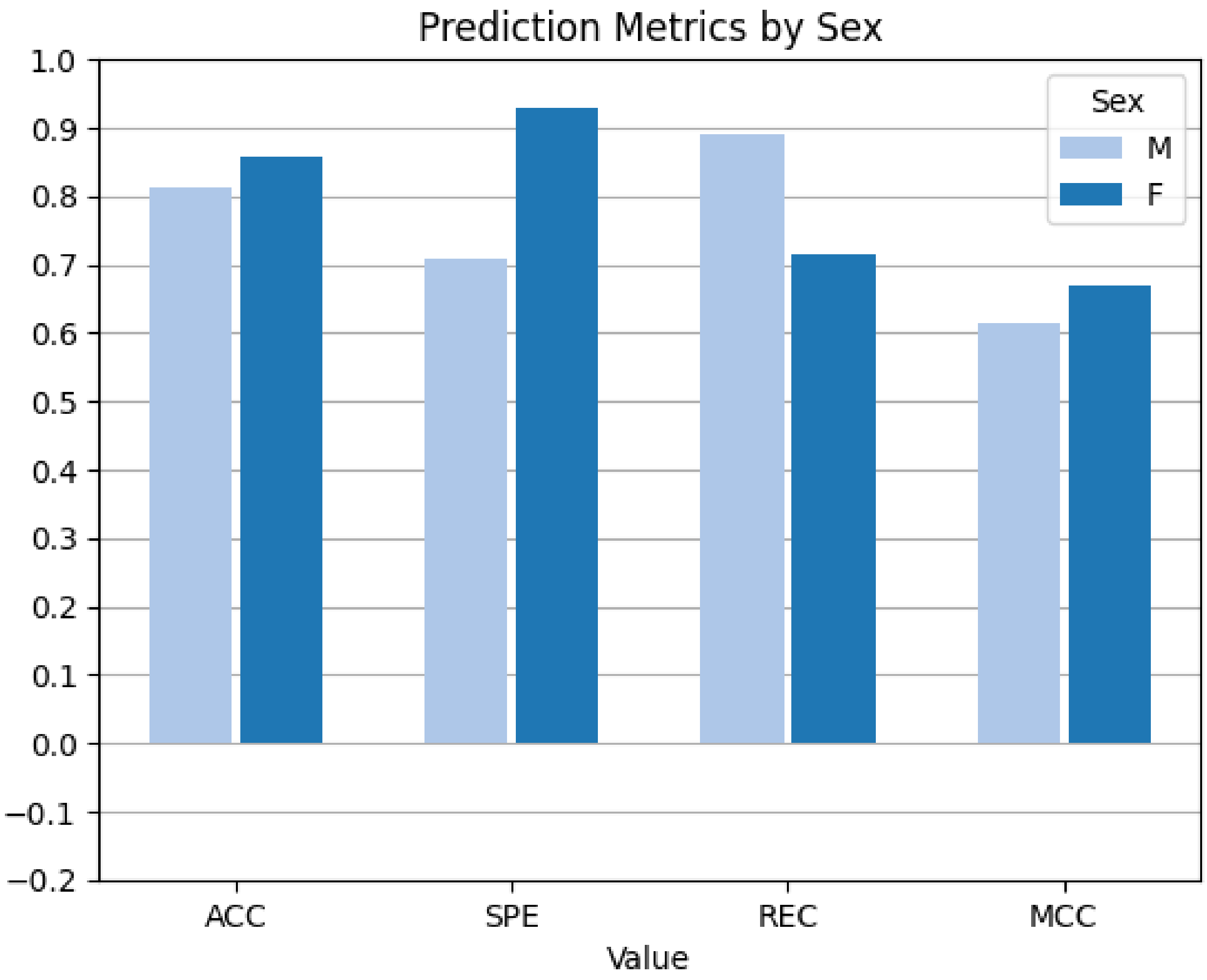
Results

Model Performance

- **Ideal value** for ACC, SPE, REC, and MCC is **1.0**
- ACC, SPE, and REC **support the model being effective** at determining AD diagnosis
- MCC suggests a **moderate positive correlation** between predicted and actual diagnosis



Model Performance

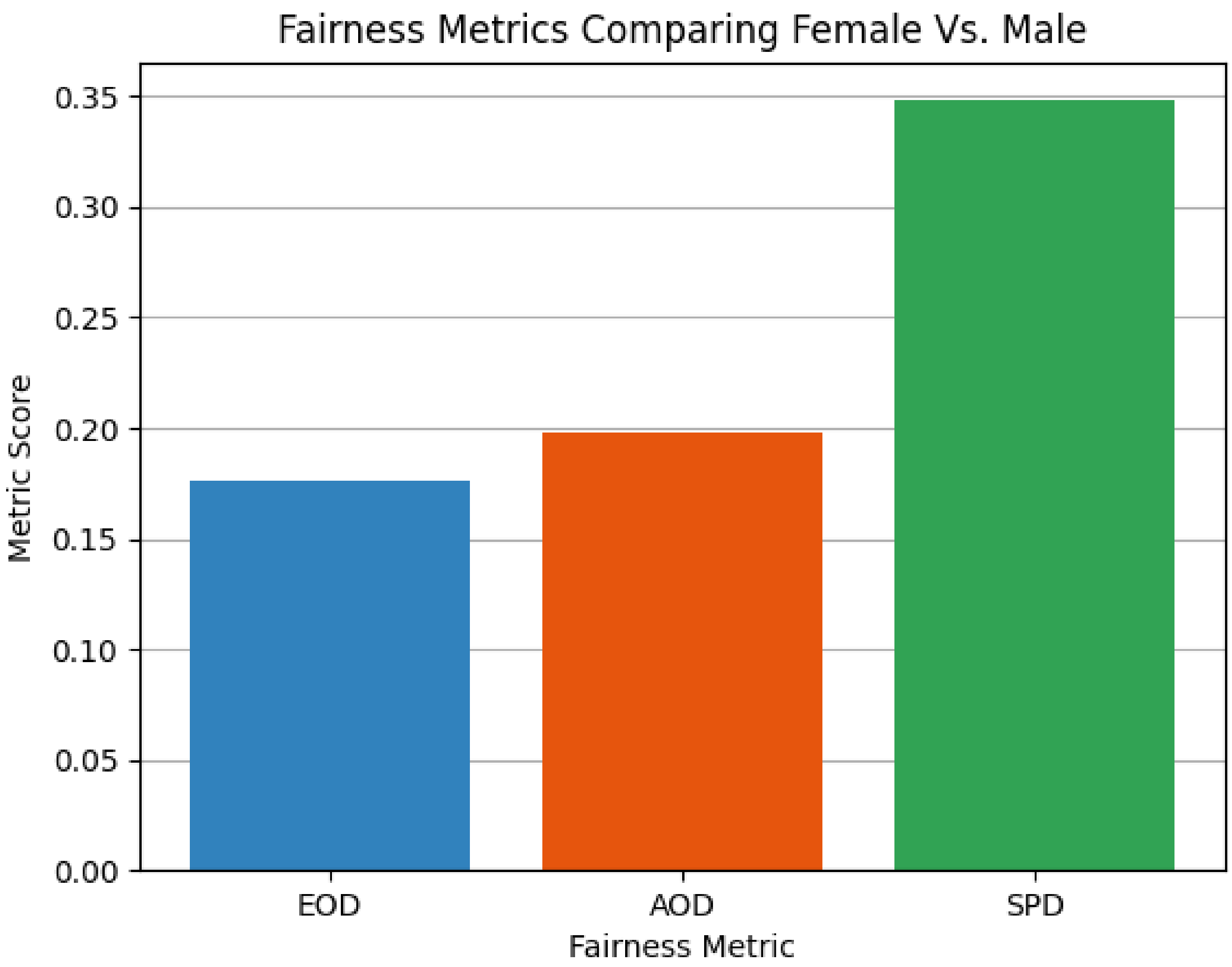


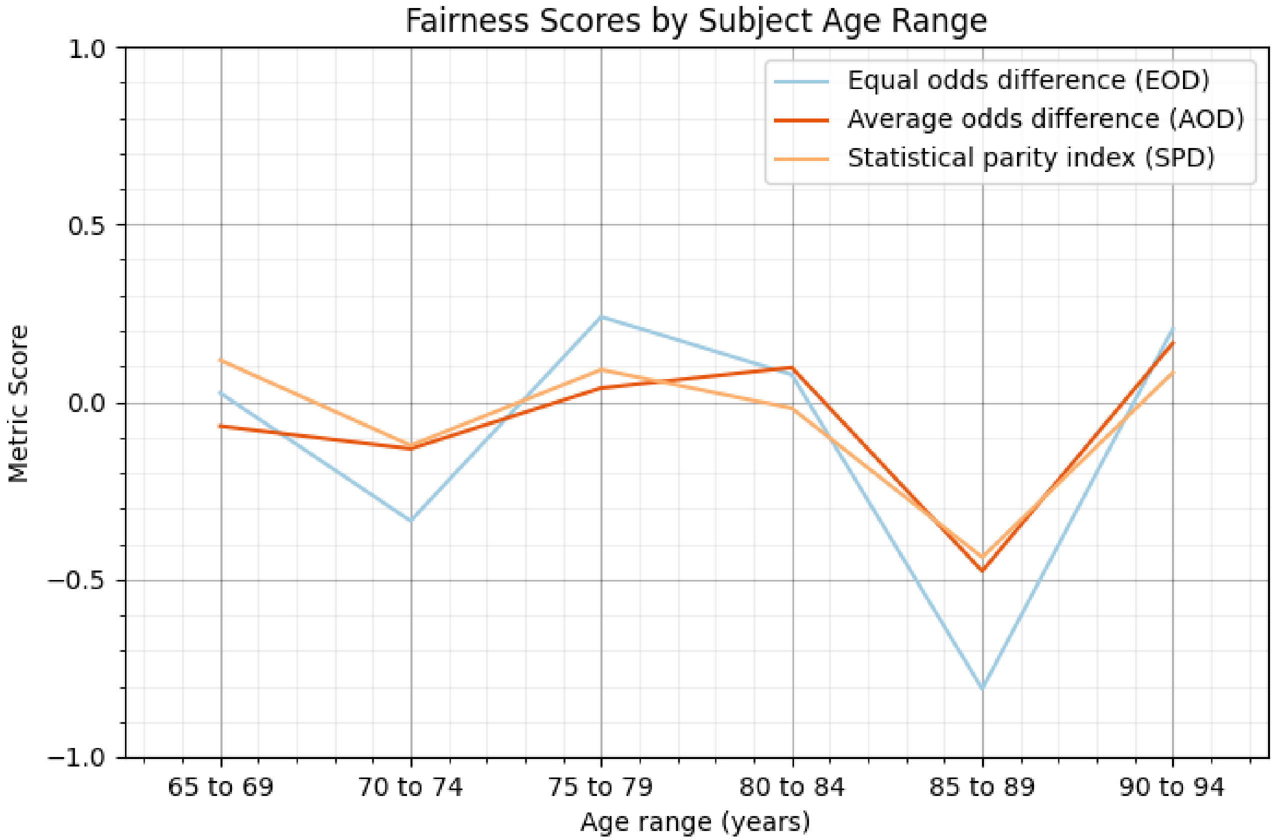
Metric	Male	Female	Difference
ACC	0.812	0.857	0.045
SPE	0.708	0.929	0.221
REC	0.891	0.714	0.177
MCC	0.615	0.671	0.056

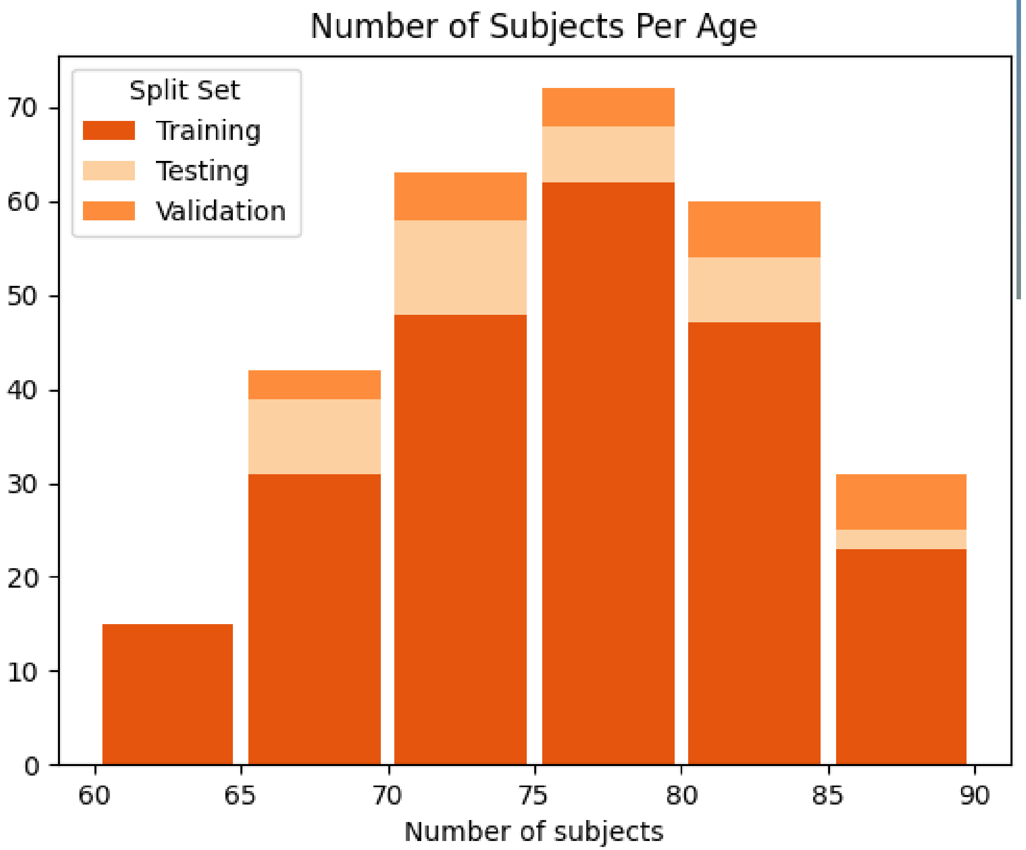
Model Fairness

Metric	Ideal Range	Model's Value	Outcome
EOD	[-0.1, 0.1]	0.176	Not ideal
AOD	[-0.1, 0.1]	0.198	Not ideal
SPD	[-0.1, 0.1]	0.348	Not ideal

- The model does **not meet** ideal fairness metrics
- Female subjects are **more likely to get a positive AD prediction** than male subjects
- The calculated inequalities are **not remarkably large**

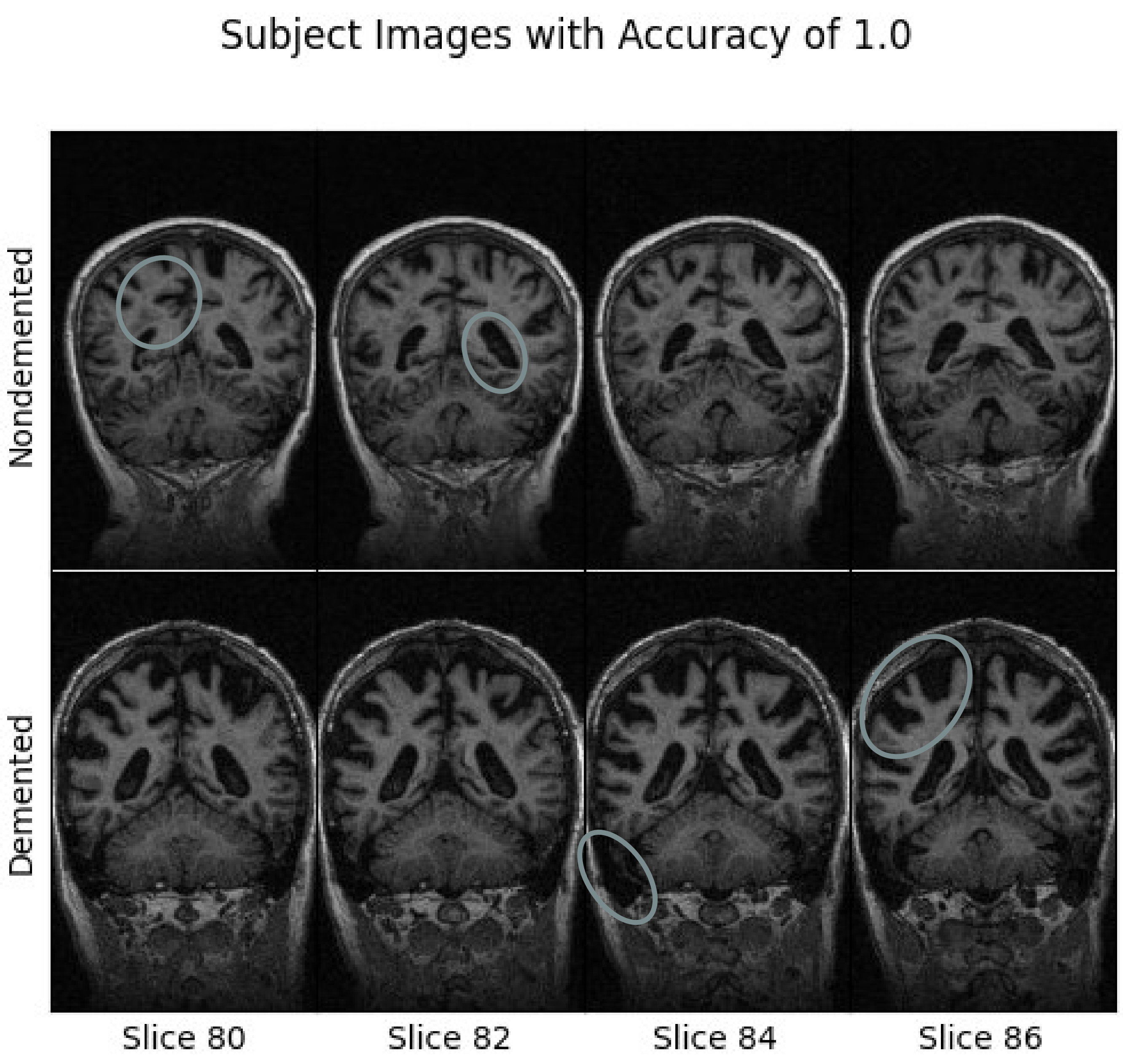



Fairness Scores by Subject Age Range
Equal odds difference (EOD)
Average odds difference (AOD)
Statistical parity index (SPD)
Metric Score
1.0
0.5
0.0
−0.5
−1.0
65 to 69
70 to 74
75 to 79
80 to 84
85 to 89
90 to 94
Age range (years)


Number of Subjects Per Age
Split Set
Training
Testing
Validation
70
60
50
40
30
20
10
0
60
65
70
75
80
85
90
Number of subjects

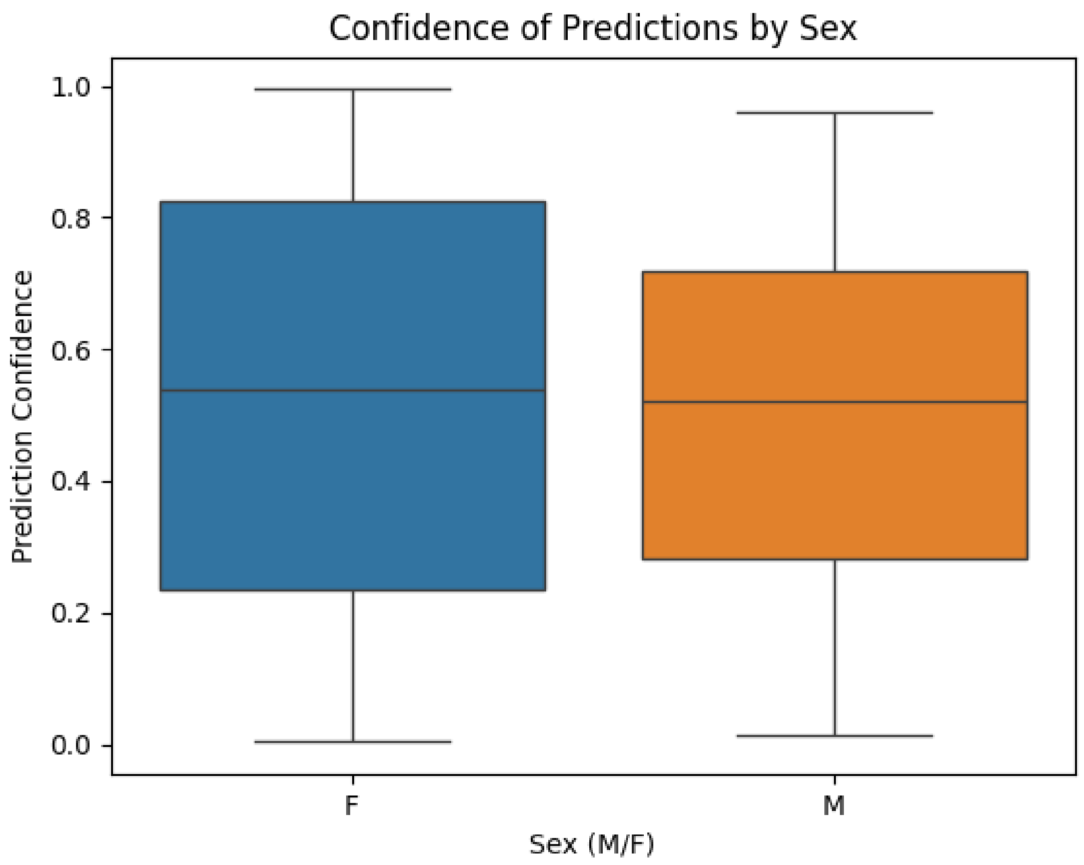
Discussion

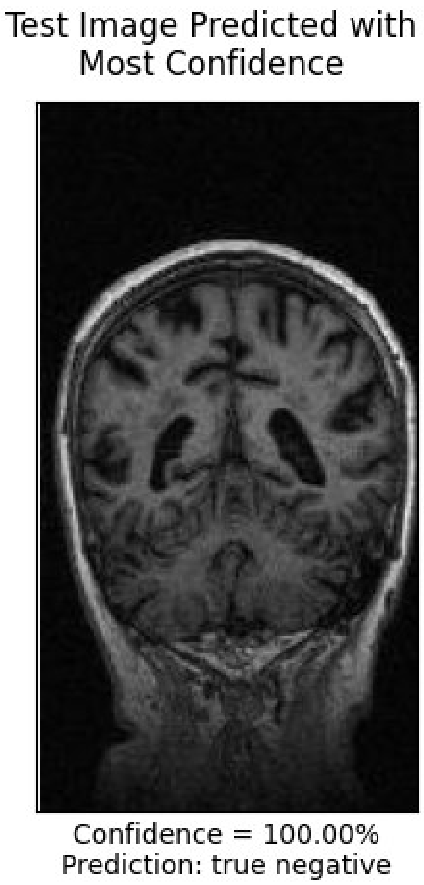
Correct Prediction of an Entire Subject



Confidence

$$\text{Confidence} = |\arg_1 - \arg_2|$$





Limitations and Challenges

Limitations

- Dataset was limited to ages 60-90, when structural changes due to typical-onset AD can appear as early as 45 years old
- Dataset only contained right-handed subjects
- Use of 2D only images requires all images to be registered to the atlas space of Talirach and Tournoux [10]

Challenges

- Overfitting of images
  - Including subjects from visits 3-5, especially those who saw AD progression, could reduce this
  - Training on fewer slices, or slices that are further apart in the brain could also help reduce overfitting

Conclusion

- OAS-52 demonstrated an acceptable performance for detecting AD in MRI acquisitions, but demonstrated bias towards patient sex

Future Work

- Investigating the use of more slices
- Externally validating OAS-52 using an external dataset
- Determining if changes to the TVT-split ratio can improve model fairness

References

[1] Alzheimer’s Association, “2024 Alzheimer’s disease facts and figures.” Alzheimer’s Association, 2024. Accessed: Mar. 06, 2025. [Online]. Available: https://www.alz.org/getmedia/76e51bb6-c003-4d84-8019-e0779d8c4e8d/alzheimers-facts-and-figures.pdf

[2] L. Xie et al., “Automated segmentation of medial temporal lobe subregions on in vivo T1-weighted MRI in early stages of Alzheimer’s disease,” Human Brain Mapping, vol. 40, no. 12, pp. 3431–3451, Aug. 2019, doi: 10.1002/hbm.24607.

[3] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, “The clinical use of structural MRI in Alzheimer disease,” Nat Rev Neurol, vol. 6, no. 2, pp. 67–77, Feb. 2010, doi: 10.1038/nrneurol.2009.215.

[4] I. Beheshti and H. Demirel, “Feature-ranking-based Alzheimer’s disease classification from structural MRI,” Magnetic Resonance Imaging, vol. 34, no. 3, pp. 252–263, Apr. 2016, doi: 10.1016/j.mri.2015.11.009.

[5] P. A. Yushkevich et al., “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” NeuroImage, vol. 31, no. 3, pp. 1116–1128, Jul. 2006, doi: 10.1016/j.neuroimage.2006.01.015.

[6] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults,” Journal of Cognitive Neuroscience, vol. 22, no. 12, pp. 2677–2684, Dec. 2010, doi: 10.1162/jocn.2009.21407.

[7] S. A. Hicks et al., “On evaluation metrics for medical applications of artificial intelligence,” Sci Rep, vol. 12, no. 1, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.

[8] F. Dehghani, N. Malik, J. Lin, S. Bayat, and M. Bento, “Fairness in Healthcare: Assessing Data Bias and Algorithmic Fairness,” in 2024 20th International Symposium on Medical Information Processing and Analysis (SIPAIM), Antigua, Guatemala: IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/SIPAIM62974.2024.10783630.

[9] IBM Watson Studio and Knowledge Catalog, “AI Governance: Evaluation Metrics.” IBM Corporation, 2024. Accessed: Apr. 01, 2025. [Online]. Available: https://www.ibm.com/docs/en/ws-and-kc?topic=metrics-average-odds-difference

[10] J. Talairach, P. Tournoux, and M. Rayport, Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging, Reprint. in Thieme classics. Stuttgart: Thieme [u.a.], 1997.

Thank you