

NeuroFM: A Neuropathology-optimized Foundation Model For Enhanced Diagnostic Accuracy In Alzheimer's Disease And Related Dementias

Ruchika Verma¹, Shrishtee Kandoi², Robina Afzal², Shengjia Chen¹, Jannes Jegminat¹, Michael W. Karlovich², Melissa Umphlett³, Timothy E. Richardson³, Kevin Clare³, Quazi Hossain³, Jorge Samanamud³, Phyllis L. Faust⁴, Elan D. Louis⁵, Ann C. McKee^{6,7}, Thor D. Stein^{6,7}, Jonathan D. Cherry⁸, Jesse Mez⁸, Dalilah D. Quintana Mora⁹, Melissa J. Nirenberg², Ruth H. Walker², Susan Morgello², Dennis W. Dickson¹⁰, Melissa E. Murray¹⁰, Carlos Cordón-Cardo¹¹, Nadejda M. Tsankova³, Jamie M. Walker³, Diana K. Dangoor³, Stephanie McQuillan³, Emma L. Thorn³, Claudia De Sanctis¹, Shuying Li¹², Thomas J. Fuchs¹, Kurt Farrell³, John F. Crary¹, Gabriele Campanella¹.

All affiliations are available in the abstract

ABSTRACT

Accurate neuropathological diagnosis of Alzheimer's disease (AD) and related dementias requires recognizing subtle morphological patterns across brain regions. This expertise is scarce and prone to interobserver variability. Current AI pathology foundation models¹ are trained primarily on non-neural, neoplastic surgical tissue, likely limiting their performance on CNS disease and leaving a gap in computational neuropathology. We introduce **NeuroFM**, a domain-specialized foundation model trained on ~1 billion tiles from 585,657 whole-slide images (80% brain tissue), using a DINOv2 self-supervised architecture with GMA² attention. NeuroFM outperformed five general-purpose pathology models on 12 of 59 downstream neuropathology tasks, establishing a new benchmark for AI-assisted dementia diagnosis.

METHODS

Data: 585,657 WSIs, ~1B tiles (80% brain, 20% general pathology), 14 disease categories, multiple institutions
Architecture: DINOv2 self-supervised, Gated MIL Attention (GMA²) aggregation, linear probing on frozen encoder features
Evaluation: 59 tasks categorized into 14 disease categories, encompassing Neoplasia, Neuroinfectious diseases, Neurodegenerative disorders, and other neuropathological conditions, benchmarked against Prov-GigaPath³, UNI⁴, UNI2, Virchow⁵, and Virchow2.

Table 1. Training data for NeuroFM

Collection	Slide (n)	Tiles (n)	Description
Mount Sinai Hospital			
Autopsy Service	6,730	343,015,028	Clinical post-mortem cases
Neurosurgical Service	9,805	217,310,111	Clinical pre-mortem cases
General Service	562,513	200,000,000	Curated collection of pathology cases
Autopsy Research	3,292	128,513,269	Research post-mortem from the NPBB
Other Research Collections			
Non-AD neurodegeneration	1,410	64,395,279	FTLD from multiple brain banks
Boston University	1,153	31,469,602	CTE and control tissues
Columbia University	754	15,232,711	Cerebellar disease and controls
Total	585,657	999,936,000	—

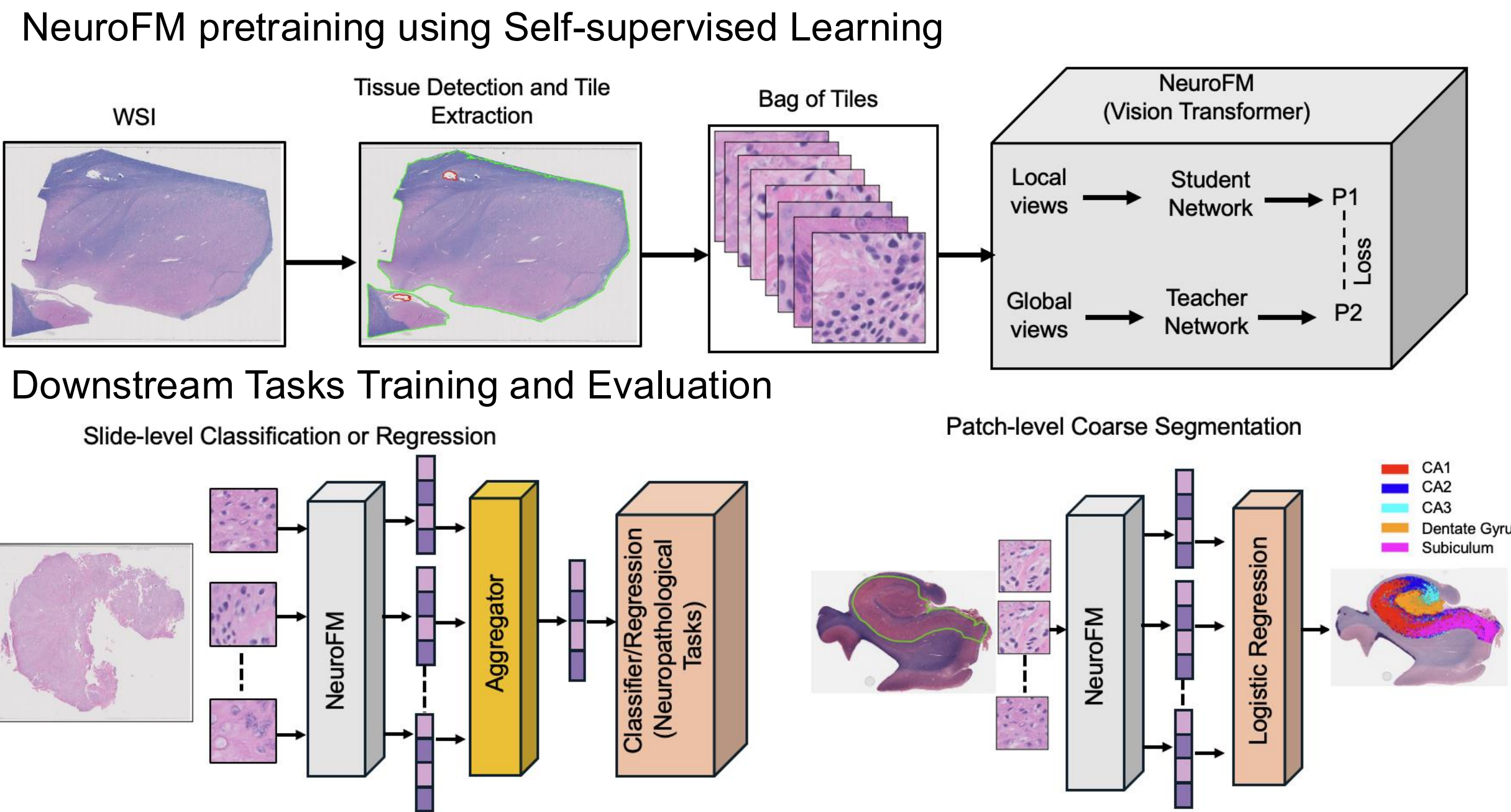


Figure 1: NeuroFM pretraining pipeline and downstream applications

RESULTS

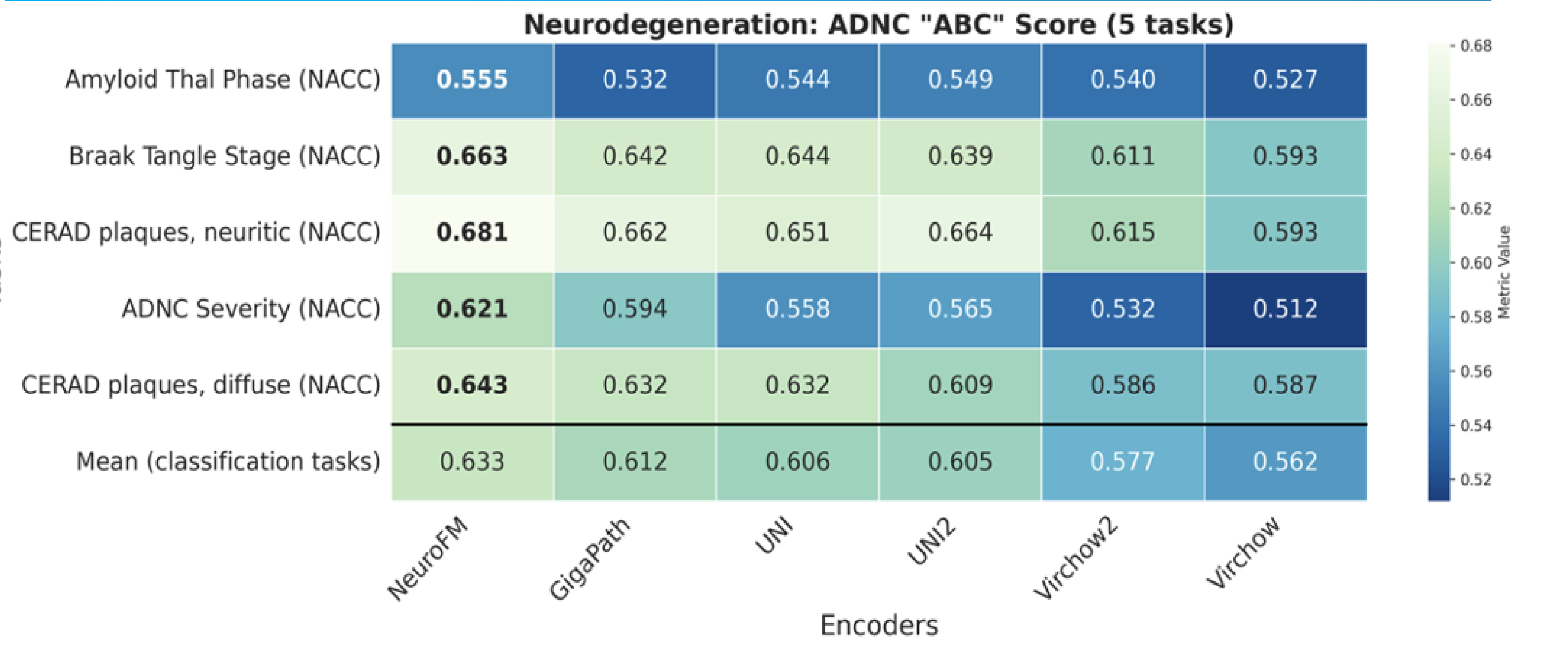


Figure 2: Heatmap of average validation AUC across encoders for five ADNC "ABC" score endpoints from >1,350 WSIs in ADRC brain banks.

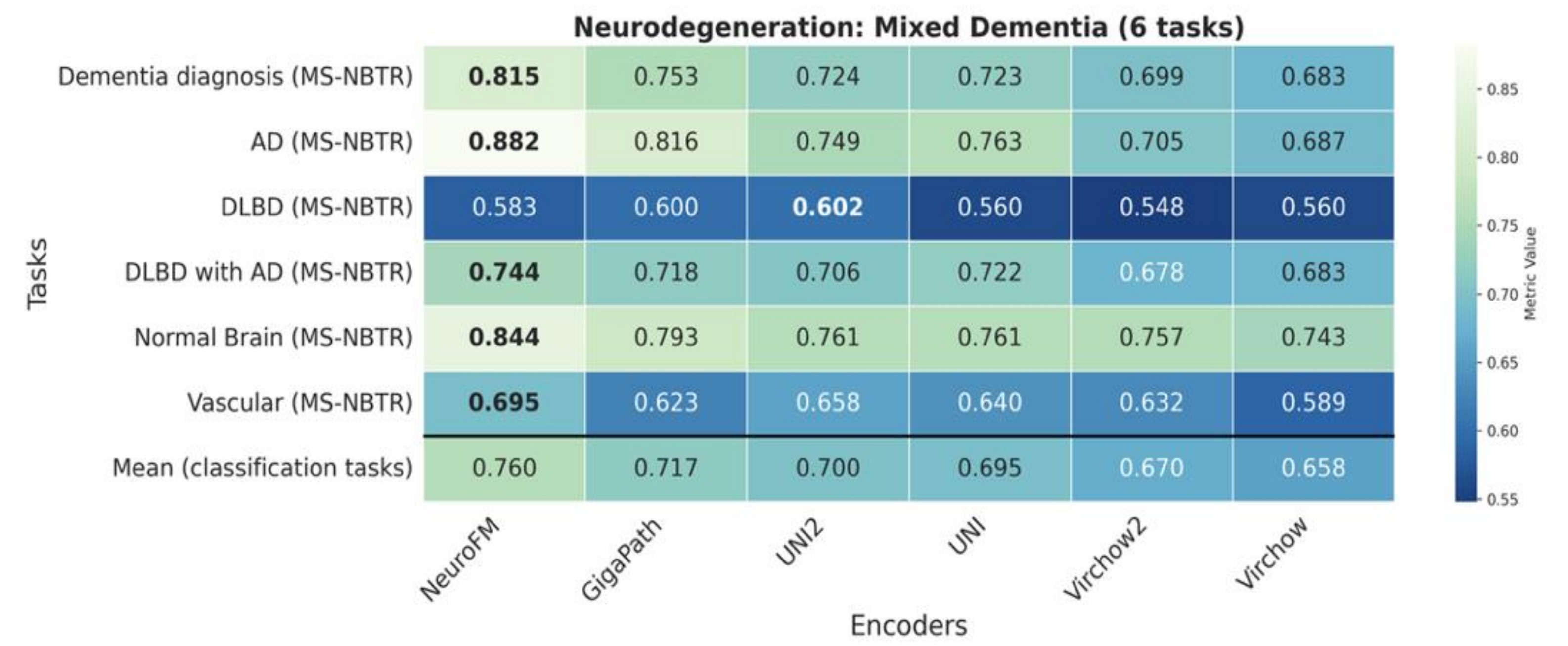


Figure 3: Heatmap of average validation AUC across encoders for mixed dementia classification tasks, evaluated on 1,704 WSIs from the Mount Sinai–NIH Brain and Tissue Repository.

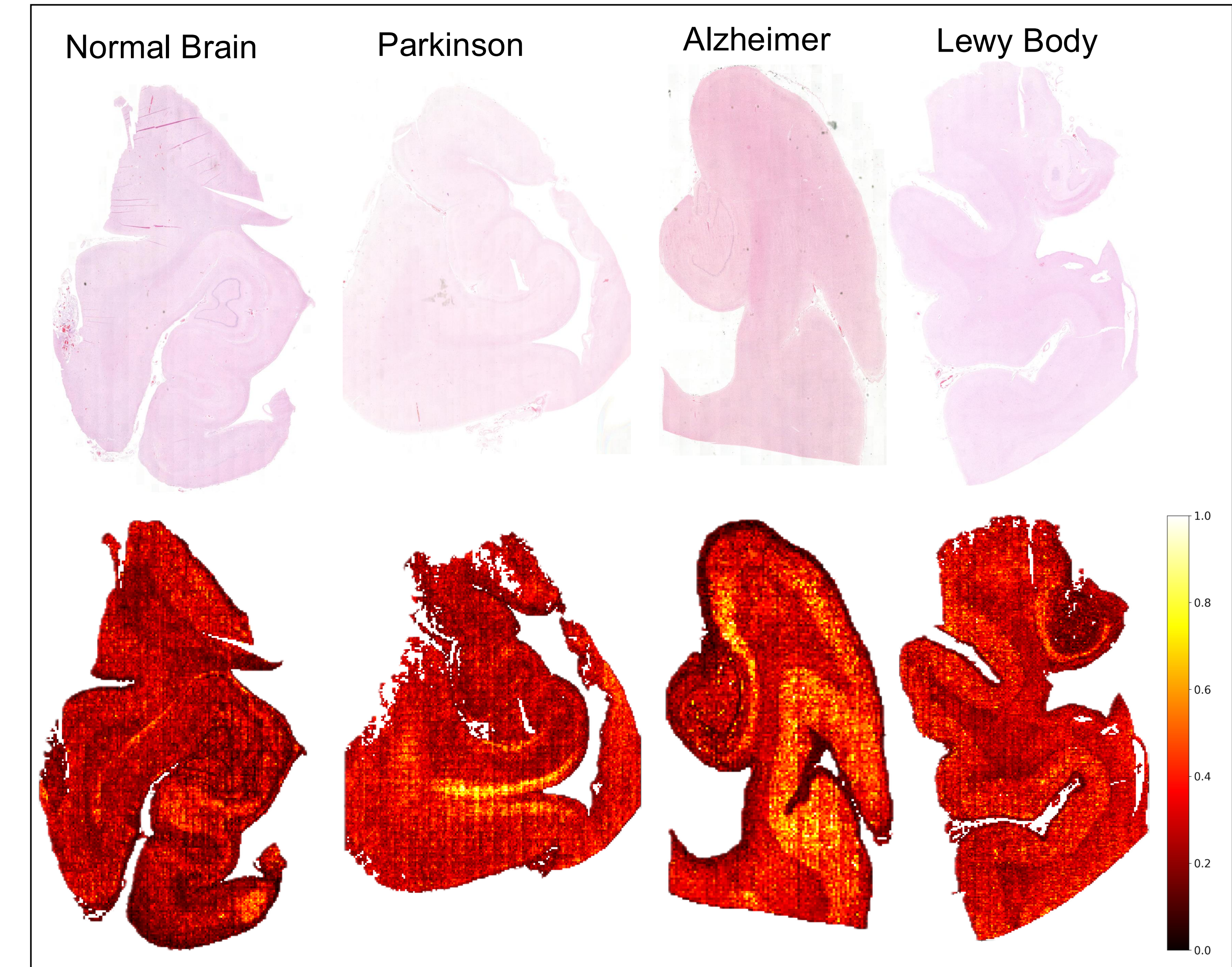


Figure 4: Representative attention maps across neurodegenerative diseases for dementia diagnosis. Top row illustrates H&E whole-slide images from Normal Brain, Parkinson disease, Alzheimer disease, and Lewy body disease cases. Bottom row represent the corresponding attention maps.



Hasso Plattner Institute for Digital Health at Mount Sinai



CONCLUSIONS

- Domain-specialized pretraining improves diagnostic accuracy for neuropathology-relevant tasks .
- NeuroFM consistently outperformed general-purpose foundation models across both ADNC "ABC" scoring and mixed dementia endpoints.
- NeuroFM demonstrate the value of brain-specific training for detecting the complex morphological patterns underlying neurodegenerative diseases and establish a framework for specialized AI tools in neuropathology.

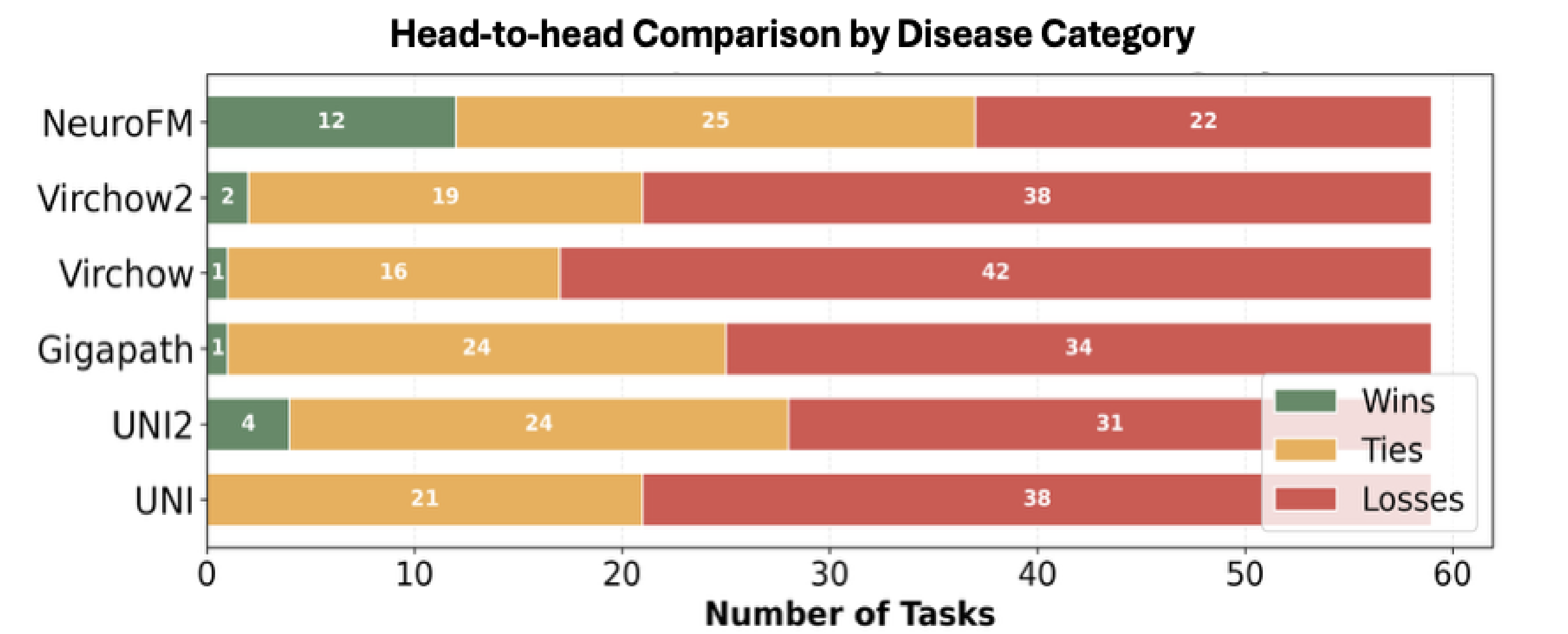


Figure 5: Task-level comparison against the top encoder. Bars show wins (green), ties (orange), and losses (red) based on statistical significance ($p < 0.05$). NeuroFM achieved the highest number of wins (12 out of 59) and fewest losses across all models, while each competitor led on only 0–4 tasks.

CLINICAL RELEVANCE

- NeuroFM improves diagnostic precision for AD and related dementias, including the mixed pathologies that are hardest to distinguish.
- It reduces interobserver variability for more consistent, reproducible assessment across institutions.
- It delivers scalable, expert-level support, extending specialized neuropathology capability where expertise is limited.
- It accelerates biomarker discovery through systematic, high-throughput analysis of large brain-tissue datasets.

ACKNOWLEDGEMENTS

We acknowledge AIR.MS, HPI.MS, and Scientific Computing and Data at ISMMS (CTSA UL1TR004419) for computational resources. This work was supported by NLM R01LM013766, VA Merit CX002342, APDA IR526-MN242.114, Neuroacanthocytosis Advocacy USA, NIA RF1AG060961, R21AG078505, RF1AG062348, R01AG061028, U19AG068753, R01AG054008, P30AG072978, P30AG066514, NINDS RF1NS095252, R01NS146414, R01NS086736, R01NS117745, RF1NS143015, R01NS142076, Manhattan HIV Brain Bank U24MH100931, Texas ARCC 957581 and 957607, and the Rainwater Charitable Foundation. WSIs from the MS-NTBR (Dr. Haroutunian; mountsinaicharcot.org), supported by the Lier Charitable Foundations. We thank V. Borukhov, Chan Zuckerberg Initiative, the NPBB Research Core (RRID: SCR_027565) staff, and all patients who donated tissue.

REFERENCES

- Moor et al. *Nature* 616.7956 (2023): 259–265.
- Ilse et al. *ICML PMLR* (2018).
- Xu et al. *Nature* (2024): 1–8.
- Chen et al. *Nat Med* 30.3 (2024): 850–862.
- Vorontsov et al. *arXiv* 2309.07778 (2023).
- Zimmermann et al. *arXiv* 2408.00738 (2024).
- Verma et al. *arXiv* 2512.05993 (2025).