IGCPHARMA

MULTIMODAL INTERPRETABLE TRANSFORMER FOR AD (MINT-AD)

Paola Ruiz Puentes[1], Néstor F. González García[1], Daniel Crovo[1], Maria C. Ibañez[1], Pablo Arbeláez[2], Ram Mukunda[1]

[1]IGC Pharma, Inc. [2]University of Los Andes, Bogotá, Colombia

POSTER #108625

TAKE-AWAY MESSAGE

Early detection and intervention are essential for altering the course of Alzheimer's disease. MINT-AD addresses this global need by combining diverse, multimodal datasets with advanced AI methods to deliver interpretable, personalized risk assessments. MINT-AD will empower general practitioners to detect Alzheimer's risk earlier and intervene before significant cognitive decline occurs.

BACKGROUND

Alzheimer's Disease (AD) affects an estimated 416 million individuals globally across the disease continuum, including those in preclinical and prodromal stages, underscoring a vast, under-recognized public health crisis[1]. Yet, early detection remains a critical unmet need, especially in primary care worldwide, including the United States, and low-resource settings where advanced diagnostics like PET scans, CSF, and blood biomarkers are expensive, invasive, and often inaccessible for patients[1]. Standard cognitive screening tools (e.g., MMSE, MoCA, ACE-III) are inconsistently applied and fail to capture the complex interplay of biological, socioeconomic, and lifestyle factors that drive AD risk. False negatives remain high (~12%)[2], meaning that many individuals with early AD pathology are missed, delaying interventions that could slow or prevent cognitive decline. The Lancet Commission estimates that up to 40% of dementia cases could be prevented or delayed by addressing modifiable risk factors like education, social engagement, and healthcare access[3]. However, general practitioners (GPs) lack tools that integrate these critical factors into actionable risk assessments, leaving the estimated 416 million individuals at risk, including millions in the U.S., without meaningful support. This proposal introduces a methodology to develop the Multimodal Interpretable Transformer for AD (MINT-AD) an AI-powered clinical support tool designed to fill this gap, enabling early risk detection and personalized intervention. (**Figure 1**).

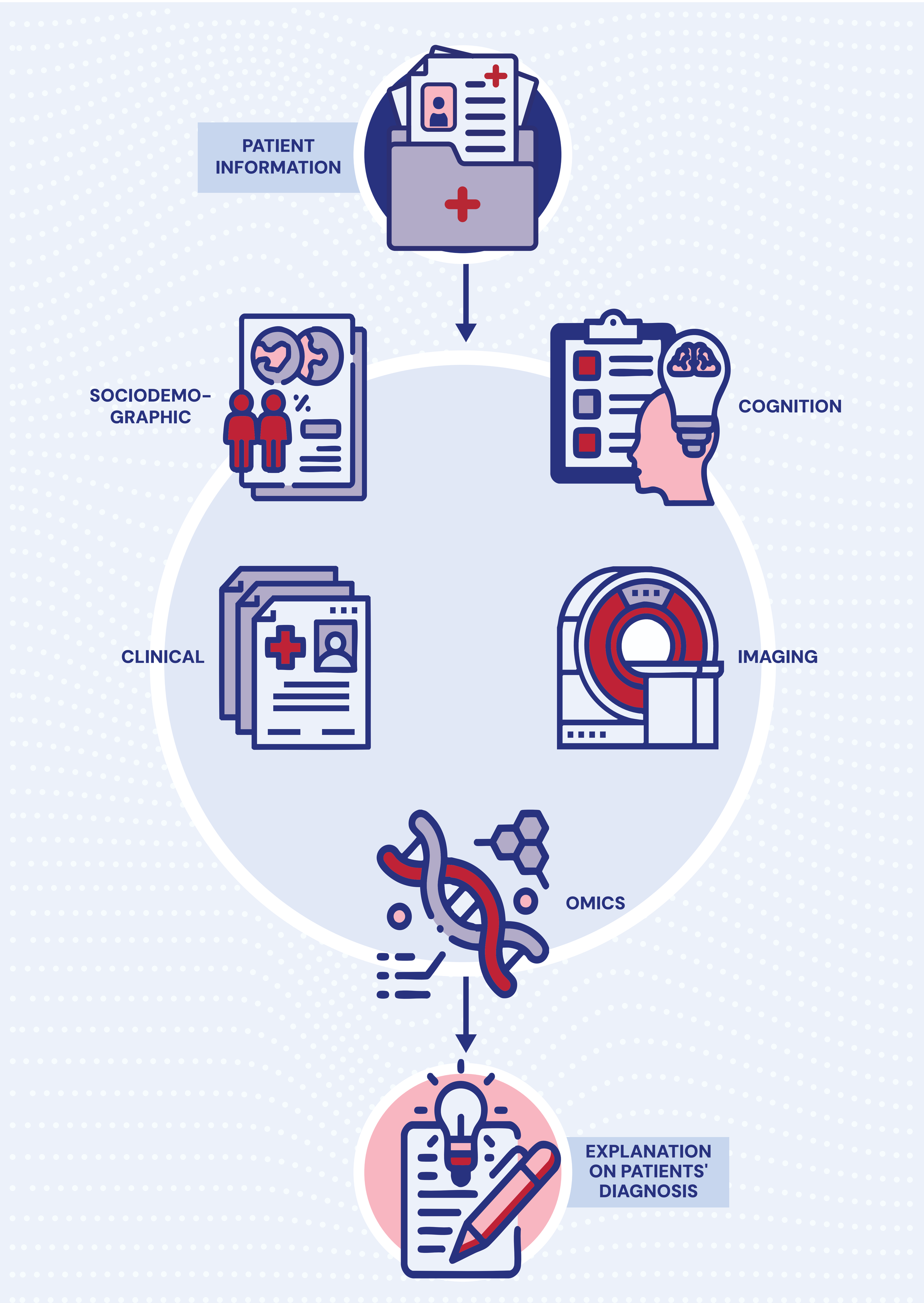


Figure 1. *Schematic view of MINT-AD.*

METHODS

Existing AI models for AD risk prediction often rely on small, homogenous datasets and focus on isolated risk factors, leading to bias and poor generalizability[2,4]. Addressing this requires models trained on diverse datasets that capture global variability in socioeconomic and health conditions. Therefore, we are developing a model trained on 32 globally diverse, multimodal datasets, with the goal of expanding to 108, that reflects the worldwide variability. These datasets include demographic, clinical, cognitive, genetic, and lifestyle information, enabling the MINT-AD to stratify AD's risk (mild, moderate, high), forecast cognitive decline two to five years in advance, and produce transparent, interpretable outputs. These outputs are designed to help GPs detect risk earlier, intervene sooner, and potentially alter the disease trajectory for millions. Our current training corpus spans six continents, incorporating 14 longitudinal studies modeled on the Health and Retirement Study (HRS) protocol with basic cognitive measures, four cognitive-ancillary cohorts, six additional aging cohorts, including three multicountry studies, one standalone cognitive study, ten clinical Alzheimer's and dementia registries, and a comprehensive genetic and epigenetic Alzheimer's dataset. (**Figure 2**).

Figure 2. *Global distribution of databases used for MINT-AD training.*

MINT-AD's architecture combines a Mixture-of-Experts framework with large language models (LLMs) trained using Chain-of-Thought prompting on patient narratives to enable intuitive interaction, alongside time-series sequence analysis for personalized cognitive decline trajectory and risk assessments.

RESULTS

Seven HRS-related databases have been harmonized by the Global2Aging initiative, with seven more in progress by our team. We have harmonized five clinical databases with ongoing efforts for the remainder. Using state-of-the-art (SOTA) transformer architectures, we have developed two early modules: one LLAMA-based[5] model predicts MMSE from sociodemographic data, and the other predicts dementia diagnosis using clinical and imaging data based on the model developed by Xue et al[6].

DISCUSSION AND CONCLUSIONS

By leveraging these diverse datasets and advanced AI models, MINT-AD aims to become a foundation model for Alzheimer's prediction and classification. Designed as a clinical support tool, it will enhance early detection, risk stratification, and intervention planning without replacing standard neuropsychological assessments. At the population level MINT-AD will enable risk mapping across largescale epidemiological variables like education, healthcare access, and environmental exposures, informing public health strategies for targeted AD prevention.

DISCLOSURE OF FUNDING

This research was funded by IGC Pharma, LLC.

REFERENCES

1. Dumas A, Destrebecq F, Esposito G, Suchonova D, Steen Frederiksen K. Rethinking the detection and diagnosis of Alzheimer's disease: Outcomes of a European Brain Council project. Aging Brain. 2023;4:100093. Epub 20230902. doi: 10.1016/j.nbas.2023.100093. PubMed PMID: 37692977; PMCID: PMC10483037.
2. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, Mitchell WG, Moukheiber L, Schirmer J, Situ J, Paguio J, Park J, Wawira JG, Yao S. Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. PLOS Digit Health. 2022;1(3):e0000022. Epub 20220331. doi: 10.1371/journal.pdig.0000022. PubMed PMID: 36812532; PMCID: PMC9931338.
3. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, Brayne C, Burns A, Cohen-Mansfield J, Cooper C, Costafreda SG, Dias A, Fox N, Gitlin LN, Howard R, Kales HC, Kivimäki M, Larson EB, Ogunniyi A, Orgeta V, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396(10248):413-46. Epub 20200730. doi: 10.1016/s0140-6736(20)30367-6. PubMed PMID: 32738937; PMCID: PMC7392084.
4. Ford E, Milne R, Curlewis K. Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. Wiley Interdiscip Rev Data Min Knowl Discov. 2023;13(3):e1492. Epub 20230219. doi: 10.1002/widm.1492. PubMed PMID: 38439952; PMCID: PMC10909482.
5. Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., … Ma, Z. (2024, November 23). The llama 3 herd of Models. arXiv.org. https://arxiv.org/abs/2407.21783
6. Xue C, Kowshik SS, Lteif D, Puducheri S, Jasodanand VH, Zhou OT, Walia AS, Guney OB, Zhang JD, Pham ST, Kaliaev A, Andreu-Arasa VC, Dwyer BC, Farris CW, Hao H, Kedar S, Mian AZ, Murman DL, O'Shea SA, Paul AB, Rohatgi S, Saint-Hilaire MH, Sartor EA, Setty BN, Small JE, Swaminathan A, Taraschenko O, Yuan J, Zhou Y, Zhu S, Karjadi C, Ang TFA, Bargal SA, Plummer BA, Poston KL, Ahangaran M, Au R, Kolachalama VB. AI-based differential diagnosis of dementia etiologies on multimodal data. medRxiv. 2024. Epub 20240326. doi: 10.1101/2024.02.08.24302531. PubMed PMID: 38585870; PMCID: PMC10996713.

SCAN ME

AI IN THE DRUG DISCOVERY PIPELINE

Néstor F. González García[1], Paola Ruiz Puentes[1], Diego Rodríguez-Soacha[1], Daniel Crovo[1], Ram Mukunda[1]

[1]IGC Pharma, Inc.

IGCPHARMA

POSTER #108629

TAKE-AWAY MESSAGE

In-silico methods powered by AI provide insights that accelerate the research and development (R&D) pipeline, reducing costs and time spent in early stages of discovery, screening, and optimization. Our in-silico screening advancements show promise into further in-silico steps and in-vitro validation assays.

BACKGROUND

Scientists spend vast amounts of resources and time to produce safe and effective treatments across a large range of diseases and symptoms. Often, 3 or more years are spent in drug discovery for compound screening and lead optimization[1]. Also, AI-driven approaches in pharma companies are increasing steadily and demonstrate potential to improve and change the drug discovery process[2]. We show a hybrid AI pipeline to accelerate the discovery process, thereby allowing researchers to reach clinical phases faster, eliminating years of lab work.

METHODS

We propose a hybrid in-silico and in-vitro pipeline to accelerate the R&D process with the following stages: in-silico compound screening; foundational models of bioactivity and contrastive learning, to identify activities of molecules, including some of ours; dynamic molecular docking for some targets of interest in AD; and finally, assays of activity to validate our in-silico findings on an in-vitro environment. **(Figure 1)**

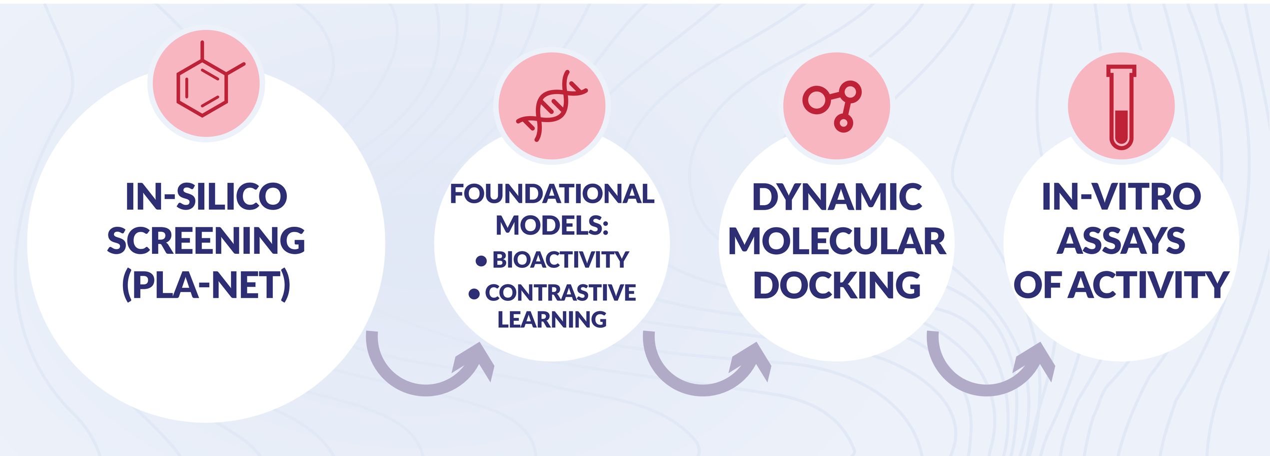


Figure 1. Hybrid in-silico and in-vitro pipeline to accelerate our R&D process.

Our advancements on this pipeline are focused on the first step. We used the protein–ligand with adversarial augmentations network (PLA-Net) to predict target–ligand interactions[3]. This model, based on a graph convolutional network, was trained on 102 protein targets, and receives a protein's and a molecule's SMILES as input to predict their interaction.

We extended the target proteins of this model with AD related targets such as dopamine and serotonin receptors and transporters, and muscarinic, oxytocin, cannabinoids, beta-adrenergic, and GLP1 receptors. We extracted molecules and their corresponding interactions with those targets from ChEMBL by filtering the target dictionary and the activities on the available assays[4]. **Figure 2** shows a summary of the tables and attributes used from the ChEMBL database schema. Then, the molecules with those activities were grouped as positive and negative examples tailored to the known function of the targets. For example, in GLP1, inhibitors and antagonists were grouped as the negative class while agonists, partial agonists, positive allosteric modulators, and activators were grouped as positive. Also, some targets had assays indicating inactivity or slight activity, which were grouped in the negative class for prediction.

Additionally, to take advantage of the previous training from PLA-Net, we looked for the best baseline model from the original training targets and used those weights to further train on each new target. We then validated the model on molecules with known activities and predicted the interaction of the targets on our patented molecules.

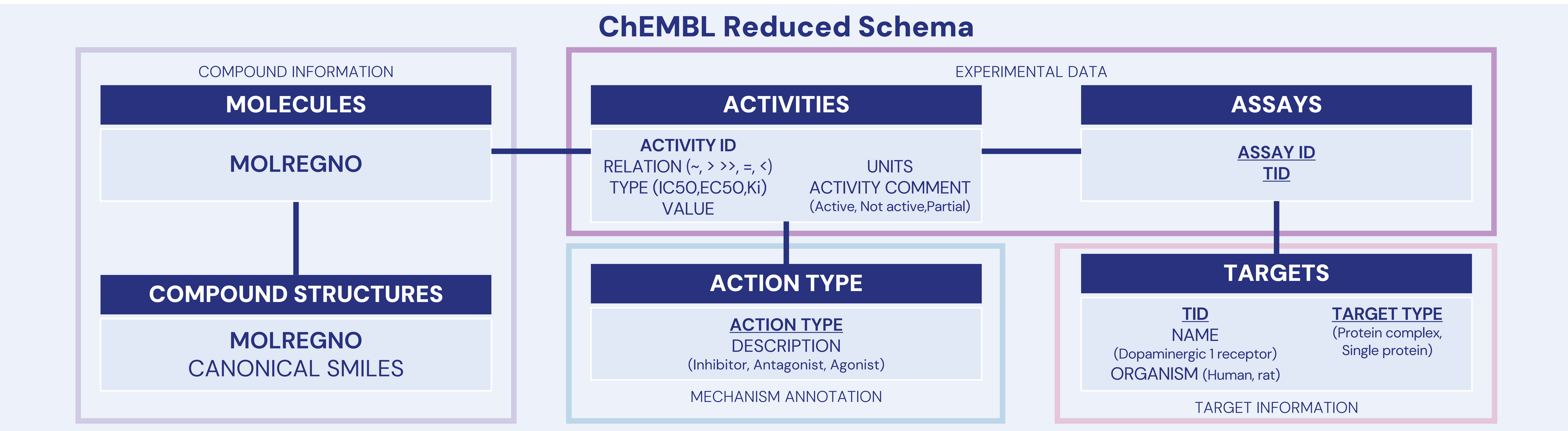


Figure 2. Subset of tables and attributes used from the ChEMBL schema to extract molecule-target bioactivity and mechanism data.

RESULTS

We successfully validated the new targets with molecules that have previously known interactions. For example, we tested Semaglutide, which is an effective GLP1 agonist, and the model predicted a 0.9316 probability of the molecule to be an agonist with the target. Table 1 shows more examples of validation of molecules known to belong to the positive class for each target. **(Table 1)**

From there, we tested 4 of our families of drugs (1's, M's, TGR's, and LMP's) on the trained targets. We identified some families with significant interactions such as the LMP's with Serotonin-1B, the TGR's with beta-adrenergic-2, and the 1's with GLP1. Table 2 has a subset of the results from this first phase. **(Table 2)**

Target	Molecule	Activity	Prediction
GLP1	Semaglutide	Agonist	0.931652
CB2	JWH-133	Agonist	0.980107
Serotonin1B	Sumatriptan	Agonist	0.701267
MuscarinicM1	Trospium	Antagonist	0.793560
MuscarinicM2			0.995641
MuscarinicM3			0.998038
MuscarinicM4			0.882812
Beta Adrenergic 2	Terbutaline	Agonist	0.991749
Dopamine Transporter	Cocaine	Inhibitor	0.960476

Table 1. Probabilities of interaction of molecules with known activity and selected targets

Molecule	GLP1	CB1	CB2	Serotonin 1B	Muscarinic M1	Muscarinic M2	Muscarinic M3	Muscarinic M4	Beta Adrenergic 2	Dopamine Transporter
1A	0.863989	0.712754	–	–	–	–	–	–	–	–
1B	0.722231	0.760048	–	–	–	–	–	–	–	–
1C	0.750051	–	–	–	–	–	–	–	–	–
M1	–	0.989819	–	–	0.724249	–	–	–	0.719102	–
M2	0.665482	–	–	–	0.730925	–	–	–	0.631194	–
M3	–	0.802999	–	–	0.684719	–	–	–	–	–
M4	0.785638	–	–	–	0.697388	–	–	–	–	–
TGR60	–	–	–	–	0.823548	0.691117	0.812315	0.807107	–	–
TGR61	–	0.628382	–	–	0.880283	0.894288	0.848915	0.725032	–	–
TGR62	–	–	–	–	–	0.772391	–	–	0.804391	–
TGR63	–	0.633334	–	–	–	–	0.664695	–	0.797183	–
TGR64	–	0.885888	–	–	–	–	–	–	0.971128	0.843516
TGR65	–	0.877609	–	–	–	–	–	–	0.936939	0.707753
LMP-01	–	–	0.872852	0.905275	–	–	–	–	0.615556	–
LMP-02	–	–	0.712781	0.897235	–	–	–	–	0.672587	–
LMP-03	–	–	–	0.937652	0.743830	–	–	–	–	–

Table 2. Probabilities of interaction between IGC's patented molecules and selected targets (only probabilities greater than 0.6)

DISCUSSION AND CONCLUSIONS

The results show the promise of an appropriate AI framework to accelerate the discovery of activity between molecules and targets, particularly in receptors with importance to therapeutic treatments for AD. For example, targeting beta-adrenergic-2 receptors because of its relation to AD's pathogenesis and amyloid beta production[5]. Moving to later stages of the proposed pipeline will highlight and validate the potential of our molecules in the field of medicine and AD.

DISCLOSURE OF FUNDING

This research was funded by IGC Pharma, LLC.

REFERENCES

1. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022 Jul;12(7):3049-3062. doi: 10.1016/j.apsb.2022.02.002. Epub 2022 Feb 11. PMID: 35865092; PMCID: PMC9293739.
2. Jayatunga, M. K., Xie, W., Ruder, L., Schulze, U., & Meier, C. (2022). AI in small-molecule drug discovery: a coming wave. Nat. Rev. Drug Discov, 21(3), 175-176.
3. Ruiz Puentes, P., Rueda-Gensini, L., Valderrama, N., Hernández, I., González, C., Daza, L., ... & Arbeláez, P. (2022). Predicting target–ligand interactions with graph convolutional networks for interpretable pharmaceutical discovery. Scientific reports, 12(1), 8434.
4. Zdrazil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., ... & Leach, A. R. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic acids research, 52(D1), D1180-D1192.
5. Yu, J. T., Wang, N. D., Ma, T., Jiang, H., Guan, J., & Tan, L. (2011). Roles of -adrenergic receptors in Alzheimer's disease: Implications for novel therapeutics. Brain Research Bulletin, 84(2), 111-117.

SCAN ME

EARLY PREDICTION OF MMSE & CSI-D SCORES WITH AI

IGCPHARMA

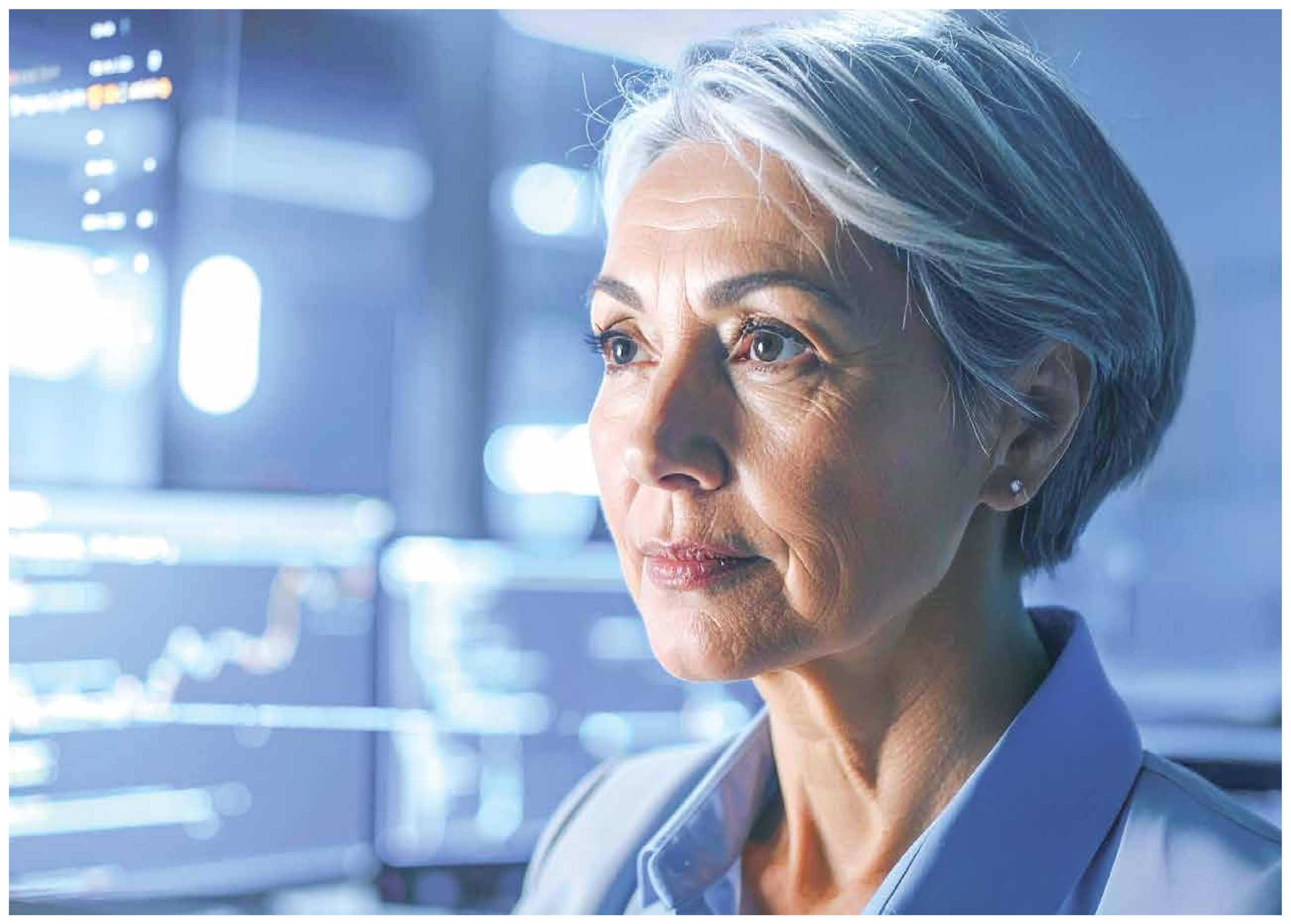
Daniel Crovo[1], Paola Ruiz Puentes[1], Néstor F. González García[1], Pablo Arbeláez[2], Ram Mukunda[1]

[1]IGC Pharma, Inc. [2]University of Los Andes, Bogotá, Colombia

POSTER #108631

TAKE-AWAY MESSAGE

Large Language Models with domain-specific pretraining demonstrate promising results for cognitive scale prediction using socioeconomic data from longitudinal studies, suggesting potential use cases in contexts where additional data such as neuroimaging or biomarkers are unavailable.



BACKGROUND

Recent studies on the application of deep learning (DL) for cognitive status prediction have demonstrated performance improvement over Machine Learning (ML) models, in particular, transformer architectures have shown promising results. For instance Aguayo et al.[1] evaluated Feedforward Neural Networks, TabTransformer, and DenseNet architectures using longitudinal ELSA[2] data to predict Parkinson's Disease, Alzheimer's Disease, dementia, and severe memory impairment, leveraging dense connectivity to uncover complex data patterns. Ma et al.[3] applied transformer-based architecture on CHARLS data to predict memory and executive functioning scores from 164 demographic, socioeconomic, and physiological features, significantly outperforming traditional SVM and XGBoost models. Zhang et al.[4] developed LSTM networks to forecast MCI risk within seven-year windows using CHARLS longitudinal data, demonstrating the effectiveness of recurrent architectures for temporal sequence modeling. Huang et al.[5] implemented a two-stage approach combining LSTM feature prediction with ensemble classifiers (GBDT, XGBoost, Random Forest) on CLHLS study data for MCI risk estimation. Gao et al.[6] conducted comprehensive comparisons between traditional ML methods (Logistic Regression, Decision Trees, XGBoost, LightGBM) and deep learning approaches (MLP, BiGRU, CNN-LSTM) using HRS data for dementia and cognitive impairment prediction, consistently showing deep learning superiority.

METHODS

We developed a data transformation pipeline to convert structured tabular data from the Mexican Health and Aging Study (MHAS)[7] into text representations describing each patient data across waves for LLM processing. Our approach used the complete MHAS dataset (26,839 data inputs across six waves, 2001-2021) for pretraining, followed by fine-tuning on the Mexican Cognitive Aging Ancillary Study (Mex-Cog)[8] subsample (n= 4,245) to predict cognitive assessment scores such as MMSE, CSID-4, CSID-6, and CSID-Informant.

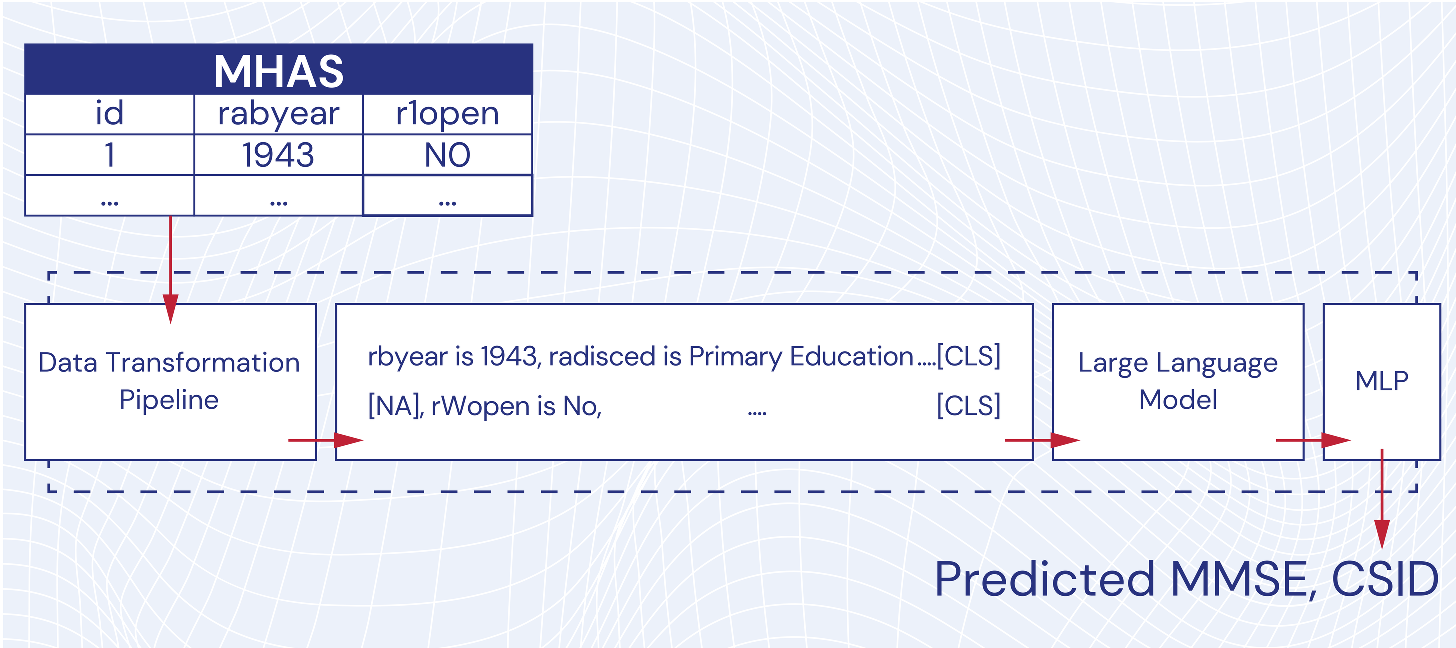


Figure 1. *Architecture for predicting cognitive scales*

The data transformation pipeline extracts the cross-wave variables such as the level of education, gender and birth year; these remain constant in the whole MHAS study. Next, it groups variable names across the five waves, since the same variable may have different names in each wave. The resulting harmonized variable names follow a common structure: rWvariable, where 'r' stands for respondent (or 's' for spouse), 'W' indicates the wave aggregation, and 'variable' is the name of the variable. For example, fall history variables from waves 3 to 5 (r3fall, r4fall, r5fall) were harmonized as rWfall.

Each variable is mapped to the value following this format: “Variable1 is value1, variable2 is value2”. We organize the mapped text representation into a sequential prompt with the cross-wave variables at the beginning followed by the ordered set of waves text representations. Each wave is separated by the token “[CLS]” enabling the model to distinguish temporal segments. Finally, we implemented a data strategy using custom tokens tailored to MHAS-specific patterns such as: "[REFUSED]" for explicit refusals, "[SKIP]" for conditional skip patterns, "[MISSING]" for missing values. This strategy allowed to use the complete patients even if there are missing values without implementing data imputation techniques or dropping the patient.

Root Mean Squared Error (RMSE) was the selected performance metric for the cognitive scales' regression task. This metric penalizes larger prediction errors more heavily and it is in the same numeric scale of the targets.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Initially we stablished a baseline performance with a Random Forest model, training only in 796 which are the patients that are in both waves of the MexCog study and have minimum missing values. Then we selected BERT[9] and LLAMA 3.2 1b[10] Large Language Models and trained them with the following two-phase strategy:

1. **Domain specific pretraining:** Using the MHAS data without including the MexCog participants, resulting in a subset of 22,594 participants, we followed standard pretraining methodologies: BERT models used Masked Language Modeling (MLM) with 15% token masking probability, while LLaMA models employed autoregressive next-token prediction.
2. **Downstream task optimization:** Following pretraining, models were fine-tuned for cognitive scales prediction using the complete Mex-Cog sample (n=4,245). We extracted fixed-length vector representations from the pretrained models: [CLS] token embeddings for BERT (768-dimensional) and final hidden state vectors for LLaMA (2048-dimensional). These representations served as inputs to a regression head comprising two linear layers (MLP). The regression head was trained using Mean Squared Error loss.

RESULTS

LLaMA models demonstrated superior performance compared to other LLMs, particularly when pretrained on MHAS data. The LLaMA model with MHAS pretraining achieved the best performance (RMSE 2.67 for MMSE), outperforming the Random Forest baseline. LLaMA models also excelled in predicting CSI-D scales, with the MHAS-pretrained version achieving the lowest RMSE across all models for CSI-D 6 (0.54) and CSI-D Informant (4.71). Even without pretraining, LLaMA outperformed BERT variants. The improved performance with pretraining demonstrates that domain-specific knowledge acquisition significantly enhances LLaMA's predictive capabilities for cognitive assessment.

Model	MMSE	CSI-D4	CSI-D6	CSI-D Informant
Random Forest	3.1439	0.9912	0.6049	6.9099
Bert without pretraining	3.6698	0.7771	0.7651	5.918
Bert with pretraining	4.5538	0.8951	0.8505	5.7247
Llama without pretraining	3.4095	0.6260	0.6786	5.3195
Llama with pretraining	2.6776	0.5709	0.5446	4.7104

Table 1. *Test set prediction results. RMSE values for the predicted scales*

DISCUSSION AND CONCLUSIONS

LLaMA was effective in predicting MMSE scores using socioeconomic data alone (RMSE=2.68) and was also effective in predicting CSI-D 6 scores (RMSE=0.54). The model can be powerful in helping doctors in resource constrained areas to help predict cognitive decline based on risk factors. Future work will focus on enhancing model interpretability to identify which socioeconomic factors most influence predictions, incorporating multiple databases and multimodal inputs and testing larger model variants to fully leverage the scaling benefits of LLMs for cognitive decline prediction across diverse populations.

DISCLOSURE OF FUNDING

This research was funded by IGC Pharma, LLC.

REFERENCES

1. G. A. Aguayo et al., “Machine learning for predicting neurodegenerative diseases in the general older population: a cohort study,” BMC Med. Res. Methodol., vol. 23, no. 1, p. 8, Jan. 2023, doi: 10.1186/s12874-023-01837-4.
2. Banks, J., Batty, G. David, Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., Zaninotto, P., “English Longitudinal Study of Ageing: Waves 0-11, 1998-2024.” UK Data Service, May 28, 2025. doi: http://doi.org/10.5255/UKDA-SN-5050-32.
3. K. Ma, J. Zhang, X. Huang, and M. Wang, “Leveraging transformer models to predict cognitive impairment: accuracy, efficiency, and interpretability,” BMC Public Health, vol. 25, no. 1, p. 504, Feb. 2025, doi: 10.1186/s12889-025-21762-z.
4. X. Zhang et al., “Establishment of a mild cognitive impairment risk model in middle-aged and older adults: a longitudinal study,” Neurol. Sci., vol. 45, no. 9, pp. 4269–4278, Sep. 2024, doi: 10.1007/s10072-024-07536-2.
5. Y. Huang et al., “Predicting mild cognitive impairment among Chinese older adults: a longitudinal study based on long short-term memory networks and machine learning,” Front. Aging Neurosci., vol. 15, Oct. 2023, doi: 10.3389/fnagi.2023.1283243.
6. H. Gao et al., “Early Identification of Cognitive Impairment in Community Environments Through Modeling Subtle Inconsistencies in Questionnaire Responses: Machine Learning Model Development and Validation,” JMIR Form. Res., vol. 8, no. 1, p. e54335, Nov. 2024, doi: 10.2196/54335.
7. R. Wong, A. Michaels-Obregon, and A. Palloni, “Cohort Profile: The Mexican Health and Aging Study (MHAS),” Int. J. Epidemiol., vol. 46, no. 2, p. e2, Apr. 2017, doi: 10.1093/ije/dyu263.
8. S. Mejia-Arango et al., “The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods,” Arch. Gerontol. Geriatr., vol. 91, p. 104210, Nov. 2020, doi: 10.1016/j.archger.2020.104210.
9. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” May 24, 2019, arXiv: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.
10. A. Grattafiori et al., “The Llama 3 Herd of Models,” Nov. 23, 2024, arXiv: arXiv:2407.21783. doi: 10.48550/arXiv.2407.21783.

SCAN ME