

**SYSTEMS AND METHODS FOR PREDICTING BRAIN TAU OR AMYLOID
LEVELS**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/513,839, filed July 14, 2023. This application also claims the benefit of U.S. Provisional Application No. 63/618,754 filed January 08, 2024. This application also claims the benefit of U.S. Provisional Application No. 63/625,245, filed January 25, 2024. This application also claims the benefit of U.S. Provisional Application No. 63/625,247, filed January 25, 2024. This application also claims the benefit of U.S. Provisional Application No. 63/644,301, filed May 08, 2024. This application also claims the benefit of U.S. Provisional Application No. 63/644,322, filed May 08, 2024. The patent applications identified above are incorporated herein by reference in their entireties to provide continuity of disclosure.

TECHNICAL FIELD

[0002] The present disclosure relates to training and using machine learning or statistical models to predict continuous or discrete-valued brain tau or amyloid β status in individual patients.

BACKGROUND

[0003] Conventional methods of determining brain tau or amyloid positivity in a patient can require administration of a radioactive tracer to the patient and subsequent collection of imaging data (e.g., performing a positron emission tomography (PET) scan). The additional time and costs can burden patients, while the required clinical resources (e.g., PET scanner) can burden healthcare providers. These burdens can prevent widespread screening of patients for tau or amyloid positivity, with implications for patient treatment and clinical study design and evaluation.

SUMMARY

[0004] Systems and, methods, and computer readable media are disclosed for predicting brain tau or amyloid β status. Consistent with disclosed embodiments a machine learning model can be trained to predict brain tau or amyloid β status of a patient. The machine learning model can predict the brain tau or amyloid β status of the patient based on biomarker data of the patient.

[0005] The disclosed embodiments include a system. The system can include at least one processor and at least one non-transitory computer readable medium containing instructions. When executed by the at least one processor, the instructions can cause the system to perform operations for predicting brain tau or amyloid β status. The operations can include obtaining a machine learning model trained to predict brain tau or amyloid β status of a patient from subject data of the patient. The subject data can include biomarker data of the patient. The brain tau or amyloid β status can include one or more continuous-valued brain tau or amyloid β levels and/or concern brain tau or amyloid β levels in multiple regions of the brain of the patient. The operations can further include generating a prediction of brain tau or amyloid β status of the patient by applying the subject data of the patient to the machine learning model. The operations can further include providing the prediction of brain tau or amyloid β status of the patient.

[0006] The disclosed embodiments include another system. The system can include at least one processor and at least one non-transitory computer readable medium containing instructions. When executed by the at least one processor, the instructions can cause the system to perform operations for training a machine learning model to predict brain tau or amyloid β status. The operations can include obtaining a brain tau or amyloid β status for a training patient. The operations can further include obtaining subject data for the training patient, the subject data including biomarker data. The operations can further include

generating a training sample that associates the brain tau or amyloid β status with the subject data for the training patient. The operations can further include training a machine learning model using the training sample to predict the brain tau or amyloid β status from the subject data. The operations can further include providing the trained machine learning model to enable prediction of the brain tau or amyloid β status.

[0007] The disclosed embodiments include a method of selecting a patient for treatment with an anti-tau therapy and/or an anti-amyloid therapy. The method can include identifying the patient as having or being at risk for AD, consistent with disclosed embodiments, and administering the anti-tau therapy or the anti-amyloid therapy.

[0008] The disclosed embodiments include a method of treating a patient having or suspected of having Alzheimer's Disease. The method can include identifying the patient as having or being at risk for AD, consistent with disclosed embodiments, and administering an anti-tau therapy and/or an anti-amyloid therapy.

[0009] The disclosed embodiments include a method of monitoring AD treatment efficacy. The method can include measuring an elevated level and/or number of brain regions including neurofibrillary tangles by obtaining a level of pTau217 and applying it to a system, consistent with disclosed embodiments. The method can further include administering a therapeutic agent and repeating the measurement. A reduction or delay in progression of brain regions including neurofibrillary tangles can indicate treatment efficacy.

[0010] The foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The accompanying drawings, which are incorporated in and constitute part of this disclosure, together with the description, illustrate and serve to explain the principles of various example embodiments.

[0012] **FIG. 1** depicts an exemplary platform for developing, validating, and deploying predictive models for predicting brain tau or amyloid β status, consistent with disclosed embodiments.

[0013] **FIG. 2** depicts an exemplary process for predicting brain tau or amyloid β status of a subject, consistent with disclosed embodiments.

[0014] **FIG. 3** depicts an exemplary process for treating a subject, consistent with disclosed embodiments.

[0015] **FIG. 4** illustrates a diagram of the progression of Alzheimer's Disease along the Braak stages, consistent with disclosed embodiments.

[0016] **FIGs. 5A-5C** illustrate exemplary structural brain network (SBN) hubs and modules, consistent with disclosed embodiments.

[0017] **FIG. 6** illustrates a table summarizing patient data, consistent with disclosed embodiments.

[0018] **FIG. 7** illustrates the distribution of Standard Uptake Value Ratio (SUVR) values for subjects with different tau positivity status, consistent with disclosed embodiments.

[0019] **FIG. 8** illustrates a table displaying the performance of tau positivity prediction, consistent with disclosed embodiments.

[0020] **FIGs. 9A-9C** illustrate top predictors for detecting tau positive subjects in Braak 3-6 using a stochastic gradient boosting machine model, consistent with disclosed embodiments.

[0021] **FIGs. 10A-10C** illustrate individual conditional expectation profiles of some features predictive of tau positivity, consistent with disclosed embodiments.

[0022] **FIGs. 11A-11C** illustrate heat maps displaying interactions between some features predictive of tau positivity, consistent with disclosed embodiments.

[0023] **FIGs. 12A-12B** illustrate plasma phosphorylated Tau181 (pTau181) as a function of tau positivity status, consistent with disclosed embodiments.

[0024] **FIGs. 13A-13C** illustrate predictors for differentiating tau positive subjects in Braak stages 3-4 and Braak stages 5-6 using Bayesian ordinal logistic models, consistent with disclosed embodiments.

[0025] **FIGs. 14A-14E** illustrate top MRI predictors, consistent with disclosed embodiments.

[0026] **FIGs. 15A-15B** illustrate cortical thickness in two regions as a function of tau positivity status in Braak stages 0-2, Braak 3-4, and Braak 5-6, grouped by inferior parietal cortical thickness-right (VCIPCR) thickness, consistent with disclosed embodiments.

[0027] **FIG. 16** illustrates a table displaying how adding Amyloid PET centiloid data can improve tau positivity prediction, consistent with disclosed embodiments.

[0028] **FIGs. 17A-17B** illustrate how amyloid PET levels can play a complementary role for predicting tau positivity, consistent with disclosed embodiments.

[0029] **FIG. 18A** provides a summary of key demographic, clinical, and genomic characteristics of the subjects included in the data used for constructing and validating certain brain tau status prediction models, consistent with disclosed embodiments.

[0030] **FIGs. 18B and 18C** depict nonlinear patterns in the relationship between plasma phosphorylated Tau217 (pTau217) / non-phosphorylated Tau217 (npTau217) ratio (pTau217R) and tau-PET SUVR values across select cortical regions and in the six Braak stage regions, consistent with disclosed embodiments.

[0031] **FIG. 19** depicts an overview of the prediction performance of certain brain tau status prediction models across various brain regions through cross-validation, consistent with disclosed embodiments.

[0032] **FIGs. 20A-20E** depict the relative influence of predictors in multivariate models for predicting brain tau status, consistent with disclosed embodiments.

[0033] **FIG. 21** depicts the prediction performance of a brain tau status prediction model using pTau217R observed during cross-validation extended to the validation set, consistent with disclosed embodiments.

[0034] **FIG. 22** depicts the prediction performance of a brain tau status prediction model using pTau217R and apolipoprotein E ϵ 4 (ApoE4) allelic count, consistent with disclosed embodiments.

[0035] **FIG. 23** depicts the prediction performance of a brain tau status prediction model using pTau217 concentration in place of pTau217R, consistent with disclosed embodiments.

[0036] **FIGs. 24A-24B** depict differences in SUVR prediction profiles and boundaries on reliable tau-PET SUVR using a brain tau status prediction model based on plasma pTau217R for the whole cortical grey matter (WCGM) and medial temporal lobe (MTL), consistent with disclosed embodiments.

[0037] **FIGs. 25A-25D** depict SUVR prediction profiles and boundaries on reliable tau-PET SUVR using a brain tau status prediction model based on plasma pTau217R for additional cortical regions beyond those shown in **FIGs. 24A** and **24B**, consistent with disclosed embodiments.

[0038] **FIGs. 26A to 26F** depict boundaries on reliable tau-PET SUVR using a brain tau status prediction model based on plasma pTau217R for the six Braak stage regions.

[0039] **FIG. 27** depicts a range of upper limits for reliable prediction of tau-PET SUVR using a brain tau status prediction model for Braak stage regions and selected cortical regions, consistent with disclosed embodiments.

[0040] **FIG. 28** depicts the ability of a brain tau status prediction model to identify subjects within the test set below or above two SUVR thresholds using plasma pTau217R, consistent with disclosed embodiments.

[0041] **FIGs. 29A-29D** depict ROC curves and corresponding AUROC values for predicting tau positivity and a tau-PET SUVR level of 1.5 in select cortical regions and the six Braak stage regions using a brain tau status model using plasma pTau217R, consistent with disclosed embodiments.

[0042] **FIG. 30** depicts the results of repeating the performance evaluation described with regards to **FIG. 28** and **FIGs. 29A to 29D** for each demographic and genomic subgroup in the validation set, consistent with disclosed embodiments.

[0043] **FIG. 31** depicts a significant potential reduction in PET scans achievable by employing a brain tau status prediction model using pTau217R, consistent with disclosed embodiments.

[0044] **FIG. 32A** depicts predicted amyloid-PET CL values, consistent with disclosed embodiments.

[0045] **FIG. 32B** depicts predicted tau-PET values, consistent with disclosed embodiments.

[0046] **FIG. 33** depicts a table summarizing patient characteristics in the study dataset, consistent with disclosed embodiments.

[0047] **FIG. 34** depicts performance of the SGB and BLLR models on the VC-1 and VC- 2 datasets, consistent with disclosed embodiments.

[0048] **FIG. 35A** depicts the percentage relative influence of predictors in an SGB model, consistent with disclosed embodiments.

[0049] **FIGs. 35B and 35C** depict ICE profiles showing individual subject-level and average outcomes, consistent with disclosed embodiments.

[0050] **FIGs. 36A to 36C** depict CL prediction range and observed CL values for three prediction models using different combinations of biomarker inputs, consistent with disclosed embodiments.

[0051] **FIG. 37** depicts the performance of prediction models for predicting amyloid status across a spectrum of CL thresholds, consistent with disclosed embodiments.

[0052] **FIGs. 38A to 38H** depict ROC curves and AUROC values for the different combinations of CL level and model for subjects in VC-1 and VC-2, consistent with disclosed embodiments.

[0053] **FIG. 39** depicts the predicted effects of using CL prediction models to screen subjects for confirmatory PET scans, consistent with disclosed embodiments.

[0054] **FIG. 40** depicts patient characteristics of a cohort for building the brain A β prediction model, consistent with disclosed embodiments.

[0055] **FIG. 41** further depicts the distribution of A β 42/A β 40 ratio and pTau217R values in the training cohort, consistent with disclosed embodiments.

[0056] **FIG. 42** depicts the odds ratio and significance of each predictor in the combined biomarker prediction model, consistent with disclosed embodiments.

[0057] **FIG. 43** depicts an ROC curve and AUROC values for the combined prediction model and three different CL values, consistent with disclosed embodiments.

[0058] **FIG. 44** depicts predicted amyloid β PET status broken out by demographic characteristics for a cohort of patients obtained through community-based screening, consistent with disclosed embodiments.

DETAILED DESCRIPTION

[0059] The following detailed description refers to the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the following description to refer to the same or similar parts. While several illustrative embodiments are described herein, modifications, adaptations, and other implementations are possible. For example, substitutions, additions, or modifications may be made to the components illustrated in the drawings, and the illustrative methods described herein may be modified by substituting,

reordering, removing, or adding steps to the disclosed methods. Accordingly, the following detailed description is not limited to the disclosed embodiments and examples. Instead, the proper scope is defined by the appended claims.

[0060] Alzheimer's disease (AD) is a progressive primary neurodegenerative disease, characterized by a decline in cognition and function, posing substantial challenges for patients, care providers, and healthcare systems globally. Key pathological features of Alzheimer's disease include the accumulation of amyloid- β ($A\beta$) plaques and the aggregation of tau proteins in the brain, which can lead to neuronal dysfunction, synaptic and neuronal loss, and eventual clinical decline. Such tau proteins (or tau) belong to the family of microtubule-associated proteins (MAPs), and are mainly expressed in neurons and found in the axons and dendrites. Tau proteins play an important role in the assembly of tubulin monomers into microtubules to constitute the cytoskeleton and serve as tracks for axonal transport. Tau proteins are translated from a single gene located on chromosome 17, with alternative mRNA splicing leading to the formation of 6 different central nervous system tau isoforms, of which 5 are found in the human adult brain. The isoforms differ, having either 3 (R1, R3, and R4) or 4 (R1-R4) repeat-regions in the carboxy (C)-terminal part and variable occurrence of microtubule binding region (MTBR). The amino (N)-terminal domain, which establishes links between microtubules and other parts of the cytoskeleton, or the plasma membrane, has a variable occurrence of 0, 1, or 2 inserts of 29 amino acids. The full-length tau isoform sequence is provided in PCT/US2023/081441 and incorporated herein by reference in its entirety. Likewise, sequences for therapeutic anti-amyloid and anti-tau antibodies useable with disclosed embodiments are provided in PCT/US2023/081441 and incorporated herein by reference in its entirety.

[0061] PET imaging serves as the gold standard for visualizing and quantifying Alzheimer's disease-related pathology in vivo. Through the use of PET tracers targeting amyloid- β

plaques and tau aggregates, clinicians can non-invasively assess these pathological features, aiding in diagnosis, prognosis, and treatment monitoring. In some embodiments, tau-PET can refer to tau positron emission tomography. A “tau-PET level” can be identified by a standard uptake value ratio (SUVr) as compared to a reference region as measured by tau-PET imaging. As used herein, a “tau-PET level”, “tau level in a brain”, “brain tau level”, and “tau load” are used interchangeably. As used herein, a tau-PET level refers to a measurement of a level of tau in a brain region, e.g., a temporal region, by PET. Methods for calculating tau-PET SUVr are known in the art and may include those described herein. In some embodiments, a Standard Uptake Value Ratio Quantitative analysis of tau-PET levels is completed using PMOD PNEURO Biomedical Image Quantification Software (PMOD Technologies, Zurich, Switzerland).

[0062] In the last decade, significant advancements in tau-specific PET radio-ligands, such as MK6240, have facilitated the targeted visualization of tau pathology in vivo. These radio-ligands offer the capability to quantify tau accumulation across various brain regions, providing valuable insights into the spatial and temporal evolution of tau pathology in Alzheimer's disease. In some embodiments, tau pathology can refer to pathological forms of tau, such as intracellular fibrillary tangles and components thereof, which are present in Alzheimer's disease (AD) and other neurodegenerative disorders, referred to as tauopathies. Aggregation of hyperphosphorylated tau into insoluble paired helical filaments (PHF) that accumulate in neurons to form neurofibrillary tangles (NFTs) are hallmarks of tau pathology. In AD, NFTs typically occur in a neuroanatomically characteristic pattern of increasing severity, generally defined according to the Braak stages 1 to 6, which correlate well with progressive neuronal loss and clinical decline. Extracellular tau seeds are also a pathological form of tau.

[0063] In some embodiments, a brain amyloid β level may be determined by longitudinal positron emission tomography (PET) assessment of an imaging agent uptake into the brain, e.g., an amyloid imaging agent. In some embodiments, the brain amyloid β value is a continuous-valued level. In some embodiments, a subject is determined to be amyloid-positive or amyloid-negative by evaluation of an amyloid PET imaging assessment. In some embodiments, the subject is “amyloid negative” if PET SUVR negativity is below a threshold determined for an amyloid PET tracer. In some embodiments, the amyloid PET tracer may be florbetaben (e.g., 18F-Florbetaben (Neuraceq®)), florbetapir (e.g., 18F-Florbetapir (Amyvid®)), and/or flutemetamol (e.g., 18F-Flutemetamol (Vizamyl®)). In some embodiments, the threshold for PET SUVR for an amyloid PET tracer is about 1.17, and a measurement below this threshold may indicate that the subject is “amyloid negative.”

[0064] As one of ordinary skill in the art will recognize, amyloid β levels from amyloid PET can be reported using the Centiloid method in “centiloid” units (CL). (Klunk WE et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's Dement.* 2015; 11:1–15 e1–4). The Centiloid method measures a tracer on a scale of 0 CL to 100 CL, where 0 is deemed the anchor-point and represents the mean in young healthy controls and 100 CL represents the mean amyloid burden present in subjects with mild to moderate severity dementia due to AD. (Id.) An elevated level of amyloid can be set relative to a baseline threshold in a healthy control determined according to methods known to a person of ordinary skill in the art (POSA). For example, a centiloid value of 32.5 can be used as a threshold value for “elevated amyloid,” and an “intermediate amyloid” level can refer to a centiloid value in the range of 20–32.5 CL (e.g., 30 CL). In another example, a centiloid value of 40 can be used as a threshold value for “elevated amyloid,” and an “intermediate amyloid” level can refer to a centiloid value in the range of 20–40 CL.

[0065] Further details on amyloid and tau-PET imaging are provided in WO 2024/118665, which is incorporated herein by reference in its entirety.

[0066] While PET imaging holds significant promise in assessing tau and/or amyloid β pathology in vivo, its widespread clinical adoption may be hindered by cost, limited availability of cyclotrons, limited accessibility for patients, few approved tau-PET or amyloid-PET tracers, and the need for specialized expertise for scan interpretation. In some cases, measurements of biomarkers in cerebrospinal fluid (CSF) and blood (e.g., plasma, serum, whole blood) may provide more accessible and cost-effective alternatives for assessing tau and/or amyloid β pathology. In particular, plasma biomarkers (e.g., plasma markers of tau phosphorylated at particular positions such as 217 (pTau217), or 181 (pTau181)) have emerged as promising alternatives for characterizing central nervous system pathology in Alzheimer's disease. Concentrations of pTau217 in plasma can be correlated with Alzheimer's disease diagnosis, severity, and disease advancement. The ratio of pTau217 to npTau217 (pTau217R) in plasma can reflect observed tau or amyloid β pathology in the brain.

[0067] As disclosed herein, predictive models can predict brain tau or amyloid β status in the whole brain of a subject (e.g., whole cortical gray matter, or the like), a brain region (e.g., a hub as described herein) specified in a neuroanatomical atlas (e.g., the Hammers atlas, Desikan-Killiany atlas, Harvard-Oxford atlas, Automated Anatomical Labeling atlas, Brainnetome atlas, or the like), a collection of such brain regions (e.g., a module as described herein), or another portion of interest of the brain of the subject.

[0068] In some embodiments, a brain tau status can include indications of tau tangles or tau deposits, while brain amyloid β status can include indications of amyloid β deposits. Such indications can pertain to the whole brain, brain region(s) (e.g., hub(s)), collection(s) of brain regions (e.g., module(s)), or the like. Such indications, for tau or amyloid β , can be

continuous-valued (e.g., a tau-PET or amyloid β SUVR level, an amyloid β centiloid level) or discrete valued (e.g., a binary classification representing the satisfaction of a diagnostic criterion, such as tau or amyloid β positivity; a multi-class classification indicating stages or classes of brain tau or amyloid β deposition; or the like). In some embodiments, a brain tau status can include predictions of tau-PET SUVR levels for multiple hubs or modules in the brain. In some embodiments, a brain amyloid β status can include predictions of amyloid β SUVR and/or centiloid levels for multiple hubs or modules in the brain. Such predictions can be made along the continuum of early Alzheimer's disease. In some embodiments, predictive models can predict regional brain tau or amyloid β levels in A β + early Alzheimer's disease patients. Thus, such a predictive model can reduce the need for PET scans to identify patients with varying degrees of brain tau or amyloid β accumulation. The simultaneous prediction of tau and/or amyloid β levels across multiple brain regions can provide enhanced flexibility and accessibility in patient screening and monitoring procedures for both clinical trials and real-world clinical settings.

[0069] As may be appreciated, predictive models consistent with disclosed embodiments can be employed as a patient screening tool to exclude individuals predicted to have no or low tau or amyloid β accumulation from clinical trials or therapy, thus limiting the need for costly and burdensome tau or amyloid β PET scans to patients prior to or during treatment. Such an approach would reduce the patient burden and potentially yield cost and time savings.

[0070] Predictive models consistent with disclosed embodiments can predict continuous-valued tau levels across a broad range of tau levels, and/or can predict continuous-valued amyloid β levels across a broad range of amyloid β levels. Such continuously valued predictions improve upon binary positive-negative assessments by enabling direct patient monitoring, assessment of treatment effects (e.g., whether a treatment lowers predicted tau or

amyloid β levels in the brain of a patient), assessing disease progression (e.g., whether the predicted tau or amyloid β levels are increasing over time in the brain of a patient), etc.

[0071] In some embodiments, predictive models consistent with disclosed embodiments can achieve good performance (e.g., as measured by mean square error (MSE) or root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), receiver operating characteristic (ROC) curves or area under ROC curves (AUROC), confusion matrices, sensitivity or selectivity, specificity, or area under such curves, precision and recall, F-measure, or any other suitable performance measure). In some embodiments, the output range of predictive models consistent with disclosed embodiments can span the range of observed outputs (e.g., tau-PET SUVR values obtained from patients, amyloid β SUVR values obtained from patients, amyloid β centiloid values obtained from patients, or the like). Furthermore, tau predictive models can show consistent performance across demographic factors (e.g., age, sex) and genomic factors (e.g., ApoE4 status) for whole brain and regional tau level predictions. Such models therefore improve upon predictive models that provide singular binary predictions (e.g., presence or absence of tau accumulation in a patient's brain).

[0072] In some embodiments, predictive models consistent with disclosed embodiments can support identification of patients for further testing (e.g., imaging, cognitive assessment testing), monitoring (e.g., follow-up with a clinician), and/or treatment (e.g., administration of a therapeutic agent, such as an anti-tau antibody or an anti-amyloid β antibody). Suitable therapeutic anti-amyloid β antibodies are known in the art and include lecanemab, while suitable therapeutic anti-tau antibodies are also known and include E2814. The predictive capability of the disclosed predictive models may allow physicians and caregivers to tailor optimal treatment and care strategies, which may lead to improved clinical outcomes.

[0073] Predictive models consistent with disclosed embodiments can include statistical and machine learning models suitable for identifying relationships between input data and output results. Such models can include regression models (e.g., logistic regression models; ridge, lasso, or elastic net regression models; time series regression models; or the like), support vector machine, gradient boosting machine, Bayes classifiers, neural networks, decision trees, random forests, other ensemble-based models, or other suitable statistical and machine learning models.

[0074] In some embodiments, suitable predictive models can include regularized logistic regression models and ensemble tree-based models. Regularized logistic regression models can include Bayesian elastic net models or Bayesian linear lasso regression (BLLR). Such models can use a suitable prior (e.g., a mixture double-exponential prior, such as a spike and slab mixture double-exponential prior) selected to reduce the complexity of the model, thus preventing overfitting and increasing model robustness.

[0075] Consistent with disclosed embodiments, ensemble tree-based models can include Stochastic Gradient Boosting (SGB) models, which can combine predictions from multiple decision trees to generate the final predictions. The nodes in each of the multiple decision trees can be trained using different random subsets of input features. The individual decision trees can therefore differ and potentially capture different signals from the data. In some embodiments, training an SGB model can include iteratively enhancing a series of simple decision trees over numerous iterations to yield refined predictions. This iterative process can progressively correct errors from previous models, ultimately converging towards more accurate predictions. In this manner, an SGB model can dynamically adapt to data intricacies during training, capturing nonlinearities and predictor interactions without pre-specifying or imposing restrictive assumptions on distributions or mathematical relationships.

[0076] Suitable input data for predictive models consistent with disclosed embodiments may comprise biomarker data, demographic data, genomic data, cognitive measurement data, and/or imaging data. In some embodiments, input data for a predictive model consistent with disclosed embodiments may include biomarker data and demographic data. In some embodiments, input data for a predictive model consistent with disclosed embodiments may include biomarker data, demographic data, and genomic data. In some embodiments, input data for predictive models consistent with disclosed embodiments may comprise biomarker data, demographic data, genomic data, and one or more of cognitive measurement data and imaging data. In some embodiments, input data for predictive models consistent with disclosed embodiments may not comprise cognitive measurement data or imaging data.

[0077] In some embodiments, biomarker data can indicate a biomarker level or function of biomarker level(s) (e.g., a normalized level of a biomarker, a ratio of the levels or normalized levels of two biomarkers, or the like). Such biomarker data can indicate whether a biomarker level or function of biomarker level(s) satisfies a condition (e.g., detectable or undetectable, exceeding a threshold value, or the like) or falls within a category (e.g., exhibits a level mapping to “elevated” in a low-transition-elevated schema, or the like). The biomarker data can be continuous, Boolean, or categorical-valued. In some embodiments, the biomarker data can be plasma, serum, or cerebrospinal fluid biomarker data. In some embodiments, the biomarker data can concern phosphorylated tau levels, non-phosphorylated tau levels, or functions or combination thereof. For example, the biomarker data can concern pTau181 or pTau217 levels. As an additional example, the biomarker data can concern a pTau217/npTau217 ratio (referred to herein as pTau217R). In some embodiments, the biomarker data can concern amyloid β protein levels, monomer levels, or functions or combination thereof. For example, the biomarker data can concern amyloid β 1-42 (A β 42) levels or amyloid β 1-40 (A β 40) levels. As an additional example, the biomarker data can

concern an A β 42/A β 40 ratio. Information on biomarkers, including pTau181, pTau217, A β 42, and A β 40, and how to measure such biomarkers, can be found in PCT/US2022/073576, which is incorporated by reference in its entirety. As an illustrative example, some biomarker levels (e.g., plasma pTau217, plasma npTau217, plasma A β 42, and/or plasma A β 40 levels) may be quantified through immunoprecipitation-mass spectrometry. In some embodiments, some biomarker levels (e.g., plasma pTau181, neurofilament light chain (NfL), and/or glial fibrillary acidic protein (GFAP) levels) may be quantified using a single molecule array assay. For example, amyloid levels may be measured using an immunoassay (e.g., a Quanterix™ Simoa® p-tau assay), C2N Diagnostics' mass spectrometry platform (PrecivityAD™) to quantitate levels and/or ratios, and/or mass spectrophotometry (IP/LC-MS/MS) based methods.

[0078] In some embodiments, demographic data can include demographic characteristics of a subject. As an example, demographic data may indicate one or more of age, sex, geographic region, occupation, education level, body mass index (BMI), race, national origin, or ethnicity. In some embodiments, genomic data can include ApoE4 status (e.g., ApoE4 allelic count).

[0079] In some embodiments, cognitive measurement data can be or include at least one assessment value, such as a clinical dementia sum of boxes (CDR-SB) measurement, an Alzheimer's Disease Composite Score (ADCOMS) measurement, an Alzheimer's Disease Assessment Scale (ADAS) measurement (e.g., ADAS-Cog-14), an Alzheimer's Disease Cooperative Study Activities of Daily Living (ADCS-ADL) score, a Mini-Mental State Examination (MMSE) score, an integrated Alzheimer's Disease rating scale (iADRS) score, a Wechsler Memory Scale-IV Logical Memory (subscale) I (WMS-IV LMI) score, or a Wechsler Memory Scale-IV Logical Memory (subscale) II (WMS-IV LMII) score. As may be appreciated, the cognitive measurement data can be or include a component or sub-score

of any suitable one of the foregoing assessments. Non-limiting examples of suitable components or sub-scores include delayed word recall (e.g., ADCDRL), word recall (e.g., ADCRL), ADCCMD (e.g., commands), ADCOF (e.g. naming objects), ADCCP (e.g., praxis such as constructional praxis or ideational praxis) CDR0101 (e.g., memory), orientation (e.g., time and location), registration (e.g., name of object), attention and calculation (e.g., spelling), language (e.g., conversation), word recognition, comprehension, and word finding.

[0080] In some embodiments, imaging data can comprise raw or processed MRI, PET, or CT images, or the like. In some embodiments, imaging data can comprise measurements (e.g., volume, surface area, cortical thickness, or the like) derived from such images. In some embodiments, imaging data can comprise images and/or measurements of brain regions identified as having particular predictive value (e.g., by an algorithm).

[0081] In some embodiments, a tau PET level is calculated from a global tau load from a whole brain signal. In some embodiments, a “tau PET level” can be identified in tau PET imaging by a standard uptake value ratio (SUVr or SUVR), e.g., to determine a level of neurofibrillary tangles in a brain region or whole brain. The SUVr may be a measurement of tau PET tracer uptake, e.g., in a region of a patient’s brain, as compared (e.g., normalized) to a reference region in the same patient. In some embodiments, an elevated level of tau, e.g. tau tangles, can be determined by comparing a PET level in a subject to that in a control subject who does not have AD. Methods for calculating tau PET SUVr are known in the art and may include quantitative analysis (e.g., computing) of SUVr by the PMOD PNEURO Biomedical Image Quantification Software (PMOD Technologies, Zurich, Switzerland). In some embodiments, a tau PET level is assessed with a PET tracer.

[0082] In some embodiments, brain tau status can include any indication of a degree of neurofibrillary tau tangles or tau deposits in the brain, including the presence or absence, quantity, location, or the like. In some embodiments, brain tau status can pertain to the brain

as a whole (e.g., whole cortical gray matter, or the like), or to one or more regions of the brain. In some embodiments, brain tau status concerns multiple regions of the brain. In some embodiments, the multiple regions can correspond to one or more Braak stages.

[0083] In various embodiments, the Braak staging system may be used to evaluate tau level in the brain, based on anatomical localization of tau neurofibrillary tangles or phospho-tau (e.g., Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 1991;82:239-59.), or a PET-based Braak staging system (e.g., as described by Therriault et al., *Nature Aging* volume 2, pages 526–535 (2022)). Accordingly, a Braak region may refer to the anatomical region or regions (e.g., temporal cortex that is commonly impacted in early AD) typically affected by tau aggregation at a Braak stage. For example, an early Braak region may refer to the temporal region such as medial temporal, meta- temporal, or temporal lobe, or to entorhinal cortex and/or hippocampus.

[0084] In general, early Braak stages may be characterized as follows: Braak stage I can be characterized by tau aggregation in the entorhinal cortex, Braak stage II can be characterized by tau aggregation in the hippocampus, Braak stage III may be characterized by tau aggregation in the amygdala, parahippocampal gyrus, fusiform gyrus, and lingual gyrus. Later Braak stages, e.g., Braak stages IV and V can be characterized by tau aggregation in the association cortices, and Braak stage VI can be characterized by tau aggregation in the primary sensory cortices.

[0085] In some preferred embodiments, tau-PET staging can use the following Braak categorizations: Braak I (entorhinal cortex); Braak II (hippocampus); Braak III (amygdala, parahippocampal gyrus, fusiform gyrus, and lingual gyrus); Braak IV (middle temporal, caudal anterior cingulate cortex, rostral anterior cingulate cortex, posterior cingulate cortex, isthmus cingulate cortex, insula, inferior temporal, temporal pole); Braak V (superior frontal, lateral orbitofrontal cortex, medial orbitofrontal cortex, frontal pole, caudal middle frontal,

rostral middle frontal, pars opercularis, pars orbitalis, pars triangularis, lateral occipital, parietal supramarginal, parietal inferior, superior temporal, parietal superior, precuneus, banks of the superior temporal sulcus, transverse temporal cortex); and Braak VI (pericalcarine cortex, postcentral gyrus, cuneus cortex, precentral gyrus, paracentral lobule, pericalcarine cortex, postcentral gyrus, cuneus cortex, precentral gyrus, paracentral lobule). [0086] In some embodiments, an alternative tau-PET staging according to Theriault et al., Nature Aging volume 2, pages 526–535 (2022) may be used to categorize stages and brains regions as follows: Braak stage I (transentorhinal), Braak stage II (entorhinal and hippocampus), Braak stage III (amygdala, parahippocampal gyrus, fusiform gyrus and lingual gyrus), Braak stage IV (insula, inferior temporal, lateral temporal, posterior cingulate and inferior parietal), Braak stage V (orbitofrontal, superior temporal, inferior frontal, cuneus, anterior cingulate, supramarginal gyrus, lateral occipital, precuneus, superior parietal, superior frontal and rostromedial frontal) and Braak stage VI (paracentral, postcentral, precentral and pericalcarine).

[0087] In Braak staging (e.g., tau-PET Braak staging), the categorization of stages and structures may vary slightly among methods due to natural variation among patients and/or variation in staging methods (e.g., different tau-PET tracers and analysis methods). Results of tau-PET Braak staging may differ from staging determined during an autopsy.

[0088] In some embodiments, the multiple regions comprise two or more regions specified in a neuroanatomical atlas. In some embodiments, a region of the brain may be a meta-region consisting of all or parts of more than one anatomical region. The temporal region for instance may comprise at least one part of the temporal lobe. For example, the temporal region may comprise the superior posterior part of the temporal lobe, the superior anterior part of the temporal lobe, the posterior part of the temporal lobe, the middle inferior part of the temporal lobe, and the fusiform gyrus. The temporal region may comprise these structures

from both the left and right hemispheres of the brain. In some embodiments, meta-regions of interest (ROIs) are defined as those regions of the brain that have certain properties, such as the regions having most tau deposition in AD patients or the regions where tau PET imaging differs between groups of patients (e.g., cognitively unimpaired individuals, e.g., those with normal amyloid PET as compared with cognitively unimpaired individuals with abnormal amyloid PET). Examples of a meta-ROI in a temporal region may be found in WO2024/118665 and Jack et al., *Alzheimer's Dementia* 13, 205–216 (2017), the contents of both of which are hereby incorporated by reference in their entireties.

[0089] In some embodiments, the multiple regions comprise regions as specified in a neuroanatomical atlas. In some embodiments, the multiple regions comprise at least 2 regions, 5 regions, 10 regions, 20 regions, 50 regions, 100 regions, 200 regions, 500 regions; at most 500 regions, 200 regions, 100 regions 50 regions, 20 regions, 10 regions, 5 regions, 2 regions; or a number of regions that is between any two of the preceding lower and upper bounds. In some embodiments, the multiple regions comprise combinations of such regions. For example, regions corresponding to the major cortical lobes of the brain (e.g., cortical, parietal, occipital, lateral temporal, medial temporal) and entire cortical grey matter (e.g., whole cortical grey matter or WCGM) can be combined into composite regions, yielding six such composite regions. Predictions for the composite regions corresponding to the major cortical lobes of the brain and WCGM can be evaluated in addition to predictions for composite regions corresponding to Braak stage.

[0090] In some embodiments, the predicted brain tau or amyloid β status can be a brain tau or amyloid β level, such as a predicted quantity, amount, or the like. The brain tau or amyloid β level can be continuous, Boolean, or categorical-valued. In some embodiments, the brain tau or amyloid β level is a continuous-valued level. In some embodiments, a predicted brain tau or amyloid β level may refer to the level that would be measured by a PET scan (e.g., a tau-

PET SUVR, an amyloid-PET SUVR, an amyloid-PET centiloid value), other imaging modalities (e.g., MRI or other modalities as described herein), histopathologically, or the like. Methods for calculating tau-PET or amyloid-PET SUVR levels are known in the art and may include those described herein.

[0091] In some embodiments, medical record information can include medical records, case notes, clinical trial records, requisition information (e.g., pertaining to biomarker testing), imaging data, or the results of laboratory tests. In some embodiments, the medical record information can include (or be useable to generate) patient medical data.

[0092] **FIG. 1** depicts an exemplary platform 100 for developing, validating, and deploying predictive models for predicting brain tau or amyloid β level(s) for subjects, consistent with disclosed embodiments. Platform 100 can be configured to obtain patient medical data from other systems (e.g., such as electronic health record (EHR) systems, insurance systems, imaging systems or medical laboratory systems, not shown in **FIG. 1**) or record(s) 101.

[0093] Consistent with disclosed embodiments, such patient medical data can include image data, biomarker data, genomic data, demographic data, cognitive measurement data, or the like. Platform 100 can be configured to generate datasets suitable for training predictive models using components such as extract transform load (ETL) engine 110 and dataset creation engine 115. Platform 100 can be configured to train models using training engine 120. Trained models can be used in the prediction phase by prediction engine 130. A user can interact with user device 199 to control and configure platform 100. The user can also provide subject data to, and receive predictions from, platform 100 by interacting with user device 199.

[0094] As may be appreciated, the particular arrangement of components depicted in **FIG. 1** is not intended to be limiting. Platform 100 can include additional components (e.g., additional databases, data sources, processing systems, or the like) or fewer components (e.g.,

by combining databases or processing systems). The functionality of the existing components can be combined or distributed among additional systems, without departing from the envisioned embodiments.

[0095] Components of **FIG. 1** can be implemented using one or more computing systems (e.g., a laptop, desktop, workstation, computing cluster, on-premises or off-premises cloud computing platform, or the like). For example, a computing cluster or workstation can implement ETL engine 110, dataset creation engine 115, or training engine 120. As an additional example, a desktop or laptop (e.g., user device 199 or another device) can implement prediction engine 130. As an additional example, the components of platform 100 (e.g., apart from user device 199) can be implemented using containerized services on a cloud computing platform. As may be appreciated, these examples are not intended to be limiting.

[0096] Consistent with disclosed embodiments, record(s) 101 can include one or more storage locations for data usable by platform 100 to predict brain tau or amyloid β status. In various embodiments such data can include medical record information for the subjects. In some embodiments, the medical record information contained in record(s) 101 can include (or be useable to generate) patient medical data, as described herein.

[0097] Consistent with disclosed embodiments, ETL engine 110 can be configured to obtain data in varying formats from one or more sources (e.g., record(s) 101, or the like). The disclosed embodiments are not limited to any particular format of the obtained data, or method for obtaining this data. For example, the obtained data can be or include structured data or unstructured data. ETL engine 110 can interact with the various data sources to receive or retrieve the data.

[0098] ETL engine 110 can transform the data into suitable format(s) and load the transformed data into a target component or database of platform 100. In some embodiments, transforming the data can include performing quality control processing on obtained data.

Such quality control processing can include confirming that data is usable (e.g., that the subject satisfies inclusion criteria for the model to be trained, that required input data for a subject is complete, or the like). In some embodiments, transforming the data can include processing the data into a standard format or structure. As may be appreciated, the input data obtained from record(s) 101 may not be in a suitable format for training a predictive model. Similarly, input data obtained from different ones of record(s) 101 may have different formats. Accordingly, ETL engine 110 can clean the obtained input data such that the input data, although originating from a variety of different sources, has a consistent format.

[0099] In some embodiments, ETL engine 110 can enrich image data or medical record information by generating additional data using the image data or medical record information. For example, ETL engine 110 can convert biomarker levels to scores (e.g., using population distribution information, clinical ranges, or the like), normalize image data (e.g., normalize area and volume by the intra cranial volume, or the like), or the like. In some embodiments, ETL engine 110 can remove unnecessary or unwanted variables or data from the input dataset. For example, when a medical record contains information unrelated to predicting tau or amyloid β status, ETL engine 110 can create a version of the medical record that contains only the information related to predicting tau or amyloid β status.

[0100] Consistent with disclosed embodiments, ETL engine 110 can load the transformed data into another component of platform 100, such as dataset creation engine 115 (or into a suitable data storage, from which dataset creation engine 115 can retrieve the data).

[0101] In some embodiments, dataset creation engine 115 can be configured to generate training samples or inference samples from data received from ETL engine 110. Dataset creation engine 115 can be configured to extract any necessary input data features from the transformed data received from ETL engine 110. In some embodiments, dataset creation engine 115 can generate features based on combinations of biomarkers (e.g., pTau217R,

A β 142/A β 40 ratio, or the like), determine correlations between input data (e.g., between thickness, area, or volume measurements for different brain regions, or the like), convert amyloid-PET SUVR scores to centiloid levels, identify brain regions as modules or hubs, as described herein, or perform other feature extraction.

[0102] In some embodiments, dataset creation engine 115 can be configured to accept label information provided by a user through user device 199. For example, dataset creation engine 115 can be configured to provide data (or metadata concerning the data) received from ETL engine 110 to user device 199 for display. In response, dataset creation engine 115 can receive label information (e.g., identification of a subject as having a particular brain tau or amyloid β level, tau or amyloid PET image labels, Braak stage information, or the like).

[0103] In some embodiments, dataset creation engine 115 can be configured to associate labels with training samples. For example, when predicting brain tau or amyloid β status, the data can include labels indicating measurements or determinations of brain tau or amyloid β status. In some embodiments, the prediction can concern brain tau or amyloid β levels measured using brain imaging data, such as tau-PET or amyloid-PET imaging data. In such embodiments, the labels can be or indicate the measured brain tau or amyloid β levels for the training samples. In some embodiments, the units for the prediction and the units for the labels can be the same (e.g., SUVR, centiloids, or the like). The dataset creation engine 115 can associate such labels with the input data. A training sample can then include the input data for the subject and the associated labels. As an additional example, when predicting brain tau or amyloid β status, a finding of brain tau or amyloid presence can be noted in a medical record of a subject (e.g., based on a biomarker assay or visual radiotracer read in PET image data). A training example can then include an indication of the finding of brain tau or amyloid presence in the input data for the subject. Dataset creation engine 115 can be configured to store training samples in data storage 105.

[0104] Consistent with disclosed embodiments, model storage 103 can be a storage location for models usable by components of platform 100 (e.g., training engine 120, or prediction engine 130). The disclosed embodiments are not limited to any particular implementation of model storage 103. Consistent with disclosed embodiments, model storage 103 can be implemented using one or more relational databases, object-oriented or document-oriented databases, tabular data stores, graph databases, distributed file systems, or other suitable data storage options.

[0105] Consistent with disclosed embodiments, data storage 105 can be a storage location for prepared datasets usable by training engine 120 or prediction engine 130. The disclosed embodiments are not limited to any particular implementation of data storage 105. Consistent with disclosed embodiments, data storage 105 can be implemented using one or more relational databases, object-oriented or document-oriented databases, tabular data stores, graph databases, distributed file systems, or other suitable data storage options.

[0106] Consistent with disclosed embodiments, training engine 120 can be configured to train, or create and train, models. Training engine 120 can be configured to create models (e.g., in response to a command to create a trained model of a particular type using an input dataset) or obtain existing models from model storage 103. Training engine 120 can be configured to create or train models using training datasets obtained from data storage 105. In some embodiments, training engine 120 can be configured to store trained models in model storage 103.

[0107] Consistent with disclosed embodiments, training engine 120 can include model training and model evaluation components. Training engine 120 can be configured to train a model using a model training component and then determine performance measure values for the model using a model evaluation component.

[0108] In some embodiments, training engine 120 can provide a model and a cross-validation or holdout portion of a training dataset to the model evaluation component. In some embodiments, training engine 120 can specify one or more performance measures as disclosed herein. Additionally, or alternatively, the model evaluation component can be configured with a predetermined or default set of such performance measures. In some embodiments, performance measure values can be displayed to a user through user device 199. The user may then interact through user device 199 with training engine 120 to update the model.

[0109] In some embodiments, training engine 120 can automatically update the model being trained based on the performance measure values. In various embodiments, training engine 120 can update the model being trained in response to user input provided through user device 199. Updating the model can include one or more of performing additional training (e.g., using the existing training dataset or another training dataset), modifying the model (e.g., changing the input features used by the model, changing the architecture of the model, or the like), or changing the training environment (e.g., changing training hyperparameters, changing a division of the training dataset into training, cross-validation, and holdout portions, or the like).

[0110] Consistent with disclosed embodiments, prediction engine 130 can be configured to predict the brain tau or amyloid β status for a subject using a prediction model. In some embodiments, prediction engine 130 can obtain the trained model from model storage 103. In some embodiments, prediction engine 130 can obtain patient medical data for the subject (e.g., subject data) from data storage 105. In some embodiments, prediction engine 130 can obtain the subject data from another data storage location. This alternative data storage location can be associated with another entity or user. For example, prediction engine 130 can

receive or retrieve the subject data from a healthcare system controlled by an entity distinct from the entity that controls prediction engine 130.

[0111] Consistent with disclosed embodiments, the subject data can include demographic data, cognitive measurement data, genomic data, imaging data, biomarker data, or other suitable subject data. The prediction engine 130 can apply the subject data to the trained prediction model to provide as output a predicted brain tau or amyloid β status. The output can be provided by prediction engine 130 to user device 199. In some embodiments, the output can be stored on a computing device associated with platform 100 or provided to another system.

[0112] Consistent with disclosed embodiments, user device 199 can provide a user interface for interacting with other components of platform 100. The user interface can be a graphical user interface. The user interface can enable a user to configure ETL engine 110 to extract, transform, and load data according to user specifications. The user interface can enable the user to specify how the transformed data received by dataset creation engine 115 is converted into labeled training data (or suitable patient data). In some embodiments, the user interface can enable the user to interact with dataset creation engine 115 to manually or semi-manually label or annotate the training data. In some embodiments, the user interface can enable a user to provide data or models to training engine 120 for training, or to prediction engine 130 for identification and classification.

[0113] In some embodiments, the user interface can enable a user to interact with training engine 120 to create or select a model for training, create or select a dataset for use in training the model, or select training parameters or hyperparameters. In some embodiments, the user interface can enable a user to interact with training engine 120 to display information related to training of the model (e.g., performance measure values, loss function values during training or functions thereof, or other training information). In some embodiments, the user

interface can enable a user to interact with prediction engine 130 to select a training model and patient data (e.g., a base image). In some embodiments, the user interface can enable a user to interact with prediction engine 130 to display prediction data, store prediction data on a computing device, or transmit the prediction data to another system.

[0114] Components of platform 100 can be implemented using one or more computing devices. Such computing devices can include tablets, laptops, desktops, workstations, computing clusters, or cloud computing platforms. In some embodiments, components of platform 100 can be implemented using cloud computing platforms. For example, one or more of ETL engine 110, dataset creation engine 115, training engine 120, and prediction engine 130 can be implemented on a cloud computing platform. In some embodiments, components of platform 100 can be implemented using on-premises systems. For example, record(s) 101 or user device 199 can be, or be hosted on, on-premises systems. As an additional example, model storage 103 or data storage 105 can be, or be hosted on, on-premises systems.

[0115] Components of platform 100 can communicate using any suitable method. In some embodiments, two or more components of platform 100 can be implemented as microservices or web services. Such components can communicate using messages transmitted on a computer network. The messages can be implemented using SOAP, XML, HTTP, JSON, RCP, or any other suitable format. In some embodiments, two or more components of platform 100 can be implemented as software, hardware, or combined software/hardware modules. Such components can communicate using data or instructions written to or read from a memory (e.g., a shared memory), function calls, or any other suitable communication method.

[0116] As may be appreciated, the particular structure of platform 100 is not intended to be limiting. Consistent with disclosed embodiments, any two or more of record(s) 101, model

storage 103, or data storage 105 can be combined, or hosted on the same computing device.

Consistent with disclosed embodiments, ETL engine 110 and dataset creation engine 115 can be omitted from platform 100. In such embodiments, datasets formatted and configured for use by training engine 120 or prediction engine 130 can be deposited in data storage 105 by another system or using another method. Consistent with disclosed embodiments, ETL engine 110 and dataset creation engine 115 can be combined. In such embodiments, data extraction, transformation, and loading can be combined with feature extraction, labeling, and sample creation. Consistent with disclosed embodiments, training engine 120 and prediction engine 130 can be combined.

[0117] Though shown with one user device 199, platform 100 could have multiple user devices. Different user devices could be associated with different entities or different users having different roles. For example, user device 199 could be associated with a software engineer or data scientist who is developing the prediction model, while another user device could be associated with a clinician who is using the prediction model.

[0118] User device 199 can be combined with one or more other components of platform 100. In some embodiments, user device 199 and at least one of ETL engine 110, dataset creation engine 115, training engine 120, or prediction engine 130 can be implemented by the same computing device. In various embodiments, user device 199 and at least one of model storage 103 or data storage 105 can be implemented by the same computing device.

[0119] As may be appreciated, platform 100 can be integrated into a method for treating subjects or for conducting clinical trials. Prediction engine 130 can use a trained prediction model and input data for the subject to predict brain tau or amyloid β status for the subject. The predicted brain tau or amyloid β status can be used to determine a patient treatment plan for the patient. In some embodiments, the brain tau or amyloid β status for the subject can be used as a covariate in determining a treatment effect in a clinical trial. In some embodiments,

clinical trial recruitment decisions can depend on predicted brain tau or amyloid β level(s).

For example, patients having predicted brain tau level(s) that satisfy inclusion criteria (e.g., Braak stage 3 or 4, but not earlier or later) can be eligible for inclusion in a clinical trial.

Similarly, patients having predicted amyloid β centiloid levels that satisfy inclusion criteria can be included in a clinical trial. In this manner, the patient population can be enriched with patients likely to demonstrate a benefit from the investigated treatment.

[0120] **FIG. 2** depicts a process 200 for predicting brain tau or amyloid β status of a subject, consistent with disclosed embodiments. For convenience of description, process 200 is described as being performed using platform 100. However, process 200 can also be performed at least in part using another computing system. Process 200 can be used to predict brain tau or brain amyloid β status. When predicting brain amyloid β status, medical record data for a patient may include, or be derived from, A β -SUVR levels or centiloid levels. In some such embodiments, the predictive models may output predicted A β -SUVR levels or centiloid levels, or A β categories (e.g., A β +, A β -, or the like). When predicting brain tau status, medical record data for a patient may include, or be derived from, tau-PET SUVR levels. In some such embodiments, the predictive models may output predicted tau-SUVR levels, or tau categories (e.g., tau positive, tau negative, or the like). Examples of the development of predictive brain tau models and brain amyloid β models are provided herein.

[0121] Process 200 can include a dataset creation phase, a training phase, and a prediction phase. In the dataset creation phase, components of platform 100 (e.g., ETL engine 110, dataset creation engine 115, or the like) can obtain training data from databases (e.g., record(s) 101, or the like). In the training phase, components of platform 100 (e.g., training engine 120, or the like) can create or refine predictive models for predicting brain tau status from subject data. In the prediction phase, components of platform 100 (e.g., prediction engine 130, or the like) can use prediction model(s) generated in the training phase to predict

brain tau status data from subject data. The predicted brain tau status data can be used to manage treatment for the subject, to include or exclude patients from a clinical trial, or as a covariate when analyzing data from a clinical trial.

[0122] In step 210 of process 200, components of platform 100 (e.g., ETL engine 110, dataset creation engine 115, or the like) can obtain training data, consistent with disclosed embodiments. In some embodiments, the training data can include subject data, consistent with disclosed embodiments.

[0123] In some embodiments, the subject data can include biomarker data, as described herein. In some embodiments, the subject data can include biomarker data and cognitive measurement, imaging, genomic data, or demographic data, as described herein. For example, the subject data may include biomarker data, such as a level of pTau217, and demographic data, such as the age, sex, race, and/or education level of the subjects. In some embodiments, the subject data can include biomarker data and cognitive measurement data or image data, as described herein. Such image data can include measurements (e.g., for one or more brain regions identified as hubs) and composite values for one or more clusters of brain regions identified as modules. In some embodiments, the subject data can include biomarker data and genomic data. For example, the subject data can include ApoE4 allelic count.

[0124] In some embodiments, when the subject data includes imaging data, the imaging data can concern measurements (e.g., volume, surface area, cortical thickness, or the like) of brain regions (e.g., hubs), collections of regions (e.g., modules), or other portions. In some embodiments, such portions may have been identified as having particular predictive value. For convenience, such portions are described herein as being identified by dataset creation engine 115. But these portions may also be identified by another suitable system. Dataset creation engine 115 can identify such portions using network analysis or multi-level clustering.

[0125] In some embodiments, dataset creation engine 115 can identify such hubs and modules using MEGENA, as described in U.S. Provisional Patent Application No. 63/561,285, and incorporated herein by reference, thereby reducing the dimensionality of the training subject data and improving the robustness of the trained predictive model.

[0126] In such embodiments, dataset creation engine 115 can obtain MRI regional measures for training subjects. The MRI regional measures can correspond to regions specified in a neuroanatomical atlas and may be normalized to reduce inter-subject variability and account for variance due to head size. Dataset creation engine 115 can construct a planar-filtered network graph using the MRI regional measures. The planar-filtered network graph can include nodes corresponding to brain regions and edges corresponding to relationships between the brain regions. The planar-filtered network graph can favor inclusion of more highly correlated brain regions. Dataset creation engine 115 can then perform a multi-level clustering analysis using the planar-filtered network graph. Dataset creation engine 115 can then perform a multiscale hub analysis to identify modules.

[0127] In such embodiments, dataset creation engine 115 can generate composite values for the identified modules using the MRI regional measures for these modules. In some embodiments, a first principal component can be calculated for a measurement type (e.g., volume, surface area, cortical thickness, or the like) over all brain regions included in the module. This first principal component can then be associated with the module. As may be appreciated, a principal component value can be generated for multiple types of measurements (e.g., each of volume, surface area, cortical thickness, or the like) and the results for these types of measurements can be associated with the module. As may be appreciated, the disclosed embodiments are not limited to using principal component analysis to generate values for modules. The regional values for the identified hubs and modules can

be used as input data for training the prediction models, consistent with disclosed embodiments.

[0128] In some embodiments, the training data can include can be or include measured brain tau data for the subjects, consistent with disclosed embodiments. In some embodiments, the measured brain tau data can be continuously valued data. As described herein, such continuously valued brain tau data can be or include an assessment of brain tau burden based on image data for the subjects. Such image data can be or include tau-PET image data showing uptake of a tracer, or image data acquired using another suitable image modality. The tau-PET tracer may be any suitable tracer, including but not limited to MK6240, RO948, PI2620, flortaucipir, or the like. Other examples of suitable tau-PET tracers include arylquinoline derivatives (e.g., [18F]THK5317 and [18F]THK5351), or a phenyl/pyridinyl-butadienyl-benzothiazone/benzothiazolium (PBB) derivative such as [11C]PBB3. In some embodiments, the tau-PET tracer is [18F]-RO-948, [18F]-PI- 2620, [18F]-JNJ-311, or [18F]-GTP1. In some embodiments, the brain tau data can be discrete-valued data, as described herein.

[0129] Similarly, when predicting brain amyloid β status, measured brain amyloid β data can continuously valued data. As described herein, such continuously valued brain amyloid β data can be or include amyloid-PET image data showing uptake of a tracer, or image data acquired using another suitable image modality. The A β PET tracer may be any suitable tracer, including but not limited to 18F-NAV4694, florbetaben, florbetapir, flutemetamol, or the like. In some embodiments, the brain amyloid β data can be discrete-valued data, as described herein.

[0130] In some embodiments, the brain tau or amyloid β data can comprise continuous-valued tau or amyloid β data. In some embodiments, the continuous-valued tau or amyloid β data can be metric data, such as intensity data, detected amount data (e.g., number of pixels

satisfying a detection criterion), or other metric data extracted from the image data for the subjects. For example, components of platform 100 (e.g., ETL engine 110, or the like) can determine tau-PET SUVR values, amyloid-PET SUVR values, and/or amyloid-PET centiloid values from PET image data, according to suitable methods. Alternatively, such components of platform 100 can receive tau-PET SUVR values, amyloid-PET SUVR values, and/or amyloid-PET centiloid values from another component of platform 100 (e.g., records 101, or the like), or another system.

[0131] In some embodiments, the subject data can be or include repeated measurements of tau or amyloid β data over time. For example, the repeated measurements can include Braak staging measurements over time for the subjects. The repeated measurements can be implicitly or expressly associated with elapsed times since the baseline. For example, the repeated measurements can be a vector (or matrix) of values, with each position in the vector (or column of the matrix) being implicitly associated with an elapsed time. As an additional example, the repeated measurements can be a set of tuples, each tuple including an elapsed time and a set of predicted levels for that elapsed time.

[0132] In some embodiments, the training data can concern subjects satisfying a cognitive impairment condition. The cognitive impairment condition can specify that the subjects have a diagnosis of a neurological disease, dysfunction, or injury (e.g., a diagnosis of AD, a diagnosis of MCI, a diagnosis of dementia, or the like), have certain signs (e.g., amyloid positivity, such as on a biomarker test or PET scan; a biomarker score, such as a plasma, serum, or cerebrospinal fluid pTau181, pTau217, pTau231, A β 42, or A β 40 level, score, or ratio; or the like). In some instances, the cognitive impairment condition can be an inclusion criterion of a clinical study. In some embodiments, at least some of the subjects may not satisfy the cognitive impairment condition. For example, such subjects may lack a diagnosis or documented signs of neurological disease, dysfunction, or injury. In some embodiments,

the components of platform 100 can obtain at least a portion of the training data from a database (e.g., record(s) 201 or the like) or another system. In some embodiments, the components of platform 100 can generate at least a portion of the training data.

[0133] In step 220 of process 200, components of platform 100 (e.g., training engine 120, or the like) can train a predictive model to predict brain tau status for a subject, consistent with disclosed embodiments. In some embodiments, training engine 120 can create the predictive model and then store the predictive model in model storage 103. In some embodiments, training engine 120 can obtain a predictive model from model storage 103, or another database or system, and then refine the model.

[0134] In some embodiments, training engine 120 can obtain hyperparameters for training the predictive model. The particular hyperparameters obtained can depend on the type of predictive model and the disclosed embodiments are not limited to any particular set of hyperparameters. For example, a neural network model may have hyperparameters governing layer arrangement and configuration, batch size, dropout, or the like. As an additional example, a gradient boosted model may have hyperparameters governing learning rate, number of trees, bagging fraction, tree depth, or the like.

[0135] In some embodiments, a user can interact with user device 199 to provide hyperparameters to training engine 120. In some embodiments, training engine 120 can receive or retrieve hyperparameters from another component of platform 100. In some embodiments, training engine 120 can generate suitable hyperparameters. For example, training engine 120 can be configured to conduct an iterative or adaptive search of a predetermined hyperparameter space (e.g., through training predictive models, evaluating the performance of the models, and updating the selected hyperparameters based on the performance of the models).

[0136] In some embodiments, training engine 120 can train the predictive model using the hyperparameters and the training data obtained in step 210. The disclosed embodiments are not limited to any particular code or instructions for training the model. In some embodiments, training engine 120 can be configured to evaluate the performance of multiple model designs using the same training dataset (e.g., Monte-Carlo Logistic Lasso models, Bayesian Logistic Elastic Net models, Bayesian ordinal logistic models, Bayesian linear lasso regression (BLLR), tree-based models (e.g., regularized random forest models), stochastic gradient boosting machine models, or the like). As may be appreciated, BLLR can moderate the weights of predictor variables, reducing model complexity by shrinking less significant variables toward zero. SGB can combine the predictions from multiple decision trees to generate the final predictions, and can automatically model the inherent non-linearity and interactions between predictors, without prior assumptions on the distribution or specific mathematical forms of the relationships between predictors and outcomes.

[0137] The performance of a model design can be determined using k-fold cross validation. In some embodiments, training engine 120 can be configured to evaluate the performance of the best-performing model design by dividing the training dataset into training and validation subsets. Training engine 120 can evaluate model designs by performing k-fold cross validation using the training subset. Training engine 120 can select a model design and evaluate the performance of that model design using the validation subset.

[0138] For example, when training engine 120 uses the R statistical package and the predictive model is a gradient boosted model, the following code can be used to train the model:

[0139] `library(gbm)`

[0140] `x = TrainingData[,xvar]`

[0141] `y = TrainingData[,yvar]`

```
[0142] train.data = cbind(y, x)
[0143] set.seed(263)
[0144] gbm.fit <- gbm(y ~ ., data=train.data,
  verbose = FALSE, distribution = "gaussian",
  shrinkage = 0.01,
  interaction.depth = 3,
  n.minobsinnode = 100,
  n.trees = 1000,
  cv.folds = 5,
  bag.fraction = 0.5,
  n.cores=1
)
```

[0145] Training engine 120 can execute this code to determine a predictive gradient boosted model using the training data and the given hyperparameter values.

[0146] In some embodiments, training engine 120 can be configured to evaluate the performance of multiple model designs using the same training dataset. The performance of a model design can be determined using k-fold cross validation. In some embodiments, training engine 120 can be configured to evaluate the performance of the best-performing model design by dividing the training dataset into training and validation subsets. Training engine 120 can evaluate model designs by performing k-fold cross validation using the training subset. Training engine 120 can select a model design and evaluate the performance of that model design using the validation subset. As may be appreciated, the evaluation can depend on the output predicted by the model. For example, a discrete-valued output (e.g., tau or amyloid β positivity status) could be evaluated based on AUROC, or the like (e.g., training engine 120 may select a prediction model yielding higher AUROC over a training model

yielding lower AUROC). As an additional example, a continuously-valued output (e.g., tau or amyloid β SUVR levels) could be evaluated based on RMSE, or the like (e.g., training engine 120 may select a prediction model yielding lower RMSE over a training model yielding higher RMSE).

[0147] In step 230 of process 200, components of platform 100 (e.g., ETL engine 110, dataset creation engine 115, prediction engine 130, or the like) can obtain subject data for an individual subject. In some embodiments, the individual subject can satisfy a cognitive impairment condition. For example, the individual subject and the subjects from which the training data was obtained can satisfy the same cognitive impairment condition. In some embodiments, the individual subject can satisfy a similar or equivalent cognitive impairment condition (e.g., the training subject may have had a diagnosis of AD, while the individual subject may have clinical findings suggestive of AD). In some embodiments, the components of platform 100 can obtain at least a portion of the individual subject data from a database (e.g., record(s) 101 or the like) or from another system. For example, platform 100 can be configured to accept prediction requests from other systems. In some embodiments, the components of platform 100 can generate at least a portion of the individual subject data.

[0148] In some embodiments, the subject data for the individual subject (e.g., the prediction subject data) can be the same as the subject data included in the training data (e.g., the training subject data). For example, when the training subject data includes biomarker data, the prediction subject data can include the same biomarker data. In another example, when the training subject data includes a combination of certain demographic data and biomarker data, the prediction subject data can include the same demographic data and biomarker data. In an additional example, when the training subject data includes biomarker data, demographic data, and imaging data, the prediction subject data can include the same biomarker data, demographic data, and imaging data.

[0149] As may be appreciated, obtaining the prediction subject data can include reformatting or arranging the prediction subject data to match the format or arrangement of the training subject data. Similarly, obtaining the prediction subject data can include handling missing values or erroneous values in the training subject data. Furthermore, when obtaining the training subject data includes generating certain values (e.g., generating composite values for modules), obtaining the prediction subject data can similarly include generating these values.

[0150] In step 240 of process 200, components of platform 100 (e.g., prediction engine 130, or the like) can predict brain tau or amyloid β status for a subject, consistent with disclosed embodiments. In some embodiments, prediction engine 130 can input the prediction subject data to the trained prediction model. The output of the trained prediction model can be a prediction of brain tau or amyloid β status for the subject. For example, a prediction model may predict the absence or presence of brain tau or amyloid β status in the entire brain. Additionally, or alternatively, a prediction model may predict the absence or presence of brain tau or amyloid β status in a portion of the brain, such as a brain regions or collection of brain regions corresponding to a Braak stage, or an anatomical region (e.g., Medial Temporal, Frontal, or Parietal lobes of the brain).

[0151] In some embodiments, the prediction of brain tau or amyloid β status for the subject may be a predicted brain tau or amyloid β status level for the whole brain. In some embodiments, the prediction of brain tau or amyloid β status for the subject may be a predicted brain tau or amyloid β status level for a portion of the brain, such as a brain region or collection of brain regions corresponding to a Braak stage (as described herein), or other anatomical portion. For example, the trained prediction model may predict the collective amount of tau, such as tau tangles or tau deposits, or collective amount of amyloid β deposits, in regions encompassed by a given Braak stage. The prediction can be continuous-valued

(e.g., a SUVR level) discrete-valued (e.g., high, low), Boolean (e.g., present, absent) or categorical-valued, and the prediction may be for one or more portions of the brain.

[0152] As may be appreciated, the type of prediction can depend on how the model is trained. For example, when the training data includes class-valued brain tau or amyloid β data, the output of the prediction model can be a predicted class, predicted classes, or predicted class likelihood(s). In another example, the training data include tau or amyloid β level(s) metric data (e.g., tau-PET SUVR level data, amyloid β -PET SUVR or centiloid level data), the output of the prediction model can be tau or amyloid β level(s) metric data (e.g., tau-PET SUVR level data, amyloid β -PET SUVR or centiloid level data).

[0153] Consistent with disclosed embodiments, platform 100 can provide the predicted brain tau or amyloid β status. Platform 100 can provide the predicted status to a user of platform 100 (e.g., by providing the predicted output class, class probabilities, or brain tau or amyloid β level(s) to user device 199 for display), or store the predicted brain tau or amyloid β status in a component of platform 199, provide the predicted brain tau or amyloid β status to another system (e.g., a system that provided a prediction request), or the like.

[0154] In some embodiments, the output of the trained prediction model in step 240 can be a sequence of predicted brain tau or amyloid β levels (e.g., predicted brain tau or amyloid β progression data). The predicted brain tau or amyloid β level progression data can be implicitly or expressly associated with elapsed times since the baseline. For example, the output can be a vector (or matrix) of values, with each position in the vector (or column of the matrix) being implicitly associated with an elapsed time. As an additional example, the predicted brain tau or amyloid β level progression data can be a set of tuples, each tuple including an elapsed time and a set of predicted levels for that elapsed time. It will be appreciated that the disclosed embodiments may involve predicting repeated measures in an individual (e.g., predicting multiple measures for the same individual). For example, the

predicted progression data can include predictions of Braak staging over time for the particular subject.

[0155] Consistent with disclosed embodiments, platform 100 can provide the predicted brain tau or amyloid β status data. Platform 100 can provide the predicted brain tau or amyloid β status data to a user of platform 100 (e.g., by providing the predicted brain tau or amyloid β status to user device 199 for display), store the predicted brain tau or amyloid β data in a component of platform 100, provide the brain tau or amyloid β status data to another system (e.g., a system that provided a prediction request), or the like.

[0156] As may be appreciated, the predicted brain tau or amyloid β status data may be manually, semi-automatically, or automatically assessed for an indication of progression of neurological disease, dysfunction, or injury. In some instances, for example, based on the brain tau or amyloid β status, the subject may be diagnosed with mild cognitive impairment. Similarly, the predicted brain tau or amyloid β status may provide an indication that the subject will progress to Alzheimer's disease. Platform 100 may be configured to automatically evaluate (e.g., using baseline data, predicted tau or amyloid β levels, and biomarker information, demographic information, or the like) the predicted brain tau or amyloid β status data and provide an indication of such brain tau or amyloid β status data.

[0157] In some embodiments, a subject (or patient) described as having mild Alzheimer's disease dementia, or mild AD dementia may satisfy the National Institute of Aging-Alzheimer's Association (NIA-AA) core clinical criteria for probable Alzheimer's disease dementia in McKhann, G.M. *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging – Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." *Alzheimer Dement.* 2011; 7:263-9. Also included herein are subjects who have a CDR score of 0.5 to 1.0 and a Memory

Box score of 0.5 or greater at screening and baseline and subjects that exhibit change in the score on the Wechsler Memory Scale-Revised Logical Memory subscale II (WMS-R LM II).

[0158] In some embodiments, a subject (or patient) described as having MCI due to AD – intermediate likelihood, may be identified as such in accordance with the NIA-AA core clinical criteria for mild cognitive impairment due to Alzheimer’s disease – intermediate likelihood (see McKhann *supra*). For example, a subject may be symptomatic but not demented, with evidence of brain amyloid pathology making them less heterogeneous and more similar to mild Alzheimer’s disease dementia subjects in cognitive and functional decline as measured by the ADCOMS Composite Clinical Score defined herein. Also included are subjects who have a CDR score of 0.5 and a Memory Box score of 0.5 or greater at screening and baseline. Furthermore, subjects who report a history of subjective memory decline with gradual onset and slow progression over the last 1 year before screening, which is corroborated by an informant, are also included herein. Memory decline and/or episodic memory impairment can be assessed in a subject by change in the score on the Wechsler Memory Scale-Revised Logical Memory subscale II (WMS-R LM II).

[0159] In some embodiments, a subject (or patient) described as having preclinical AD, pre-AD, or as being asymptomatic for AD, is a cognitively normal individual with intermediate or elevated levels of amyloid in the brain and can be identified by asymptomatic stages with or without memory complaints and emerging episodic memory and executive function deficits. Cognitively normal can include individuals who are CDR 0, or individuals within the normal ranges of scores for cognitive measurement, as described herein. Preclinical AD occurs prior to significant irreversible neurodegeneration and cognitive impairment and is typically characterized by the appearance of *in vivo* molecular biomarkers of AD and the absence clinical symptoms. Preclinical AD biomarkers that may suggest the future development of Alzheimer’s disease include, but are not limited to, one or more of

intermediate or elevated levels of amyloid in the brain by amyloid PET (e.g., a centiloid measure of about 20-40, e.g., a centiloid measure of about 20-32), fluorodeoxyglucose (FDG) PET, or tau positron emission tomography (PET), cerebrospinal fluid level of A β 1-42 and/or A β 1-42/1-40 ratio, cerebrospinal fluid level of total tau, cerebrospinal fluid level of microtubule binding region (MTBR)-tau, cerebrospinal fluid level of neurogranin, cerebrospinal fluid level of neurofilament light chain (NfL), and blood biomarkers as measured in the serum or plasma (e.g. levels of A β 1-42, the ratio of two forms of amyloid- β peptide (A β 1-42/1-40 ratio, e.g., a ratio of between about 0.092-0.094 or below about 0.092), plasma levels of plasma total tau (T-tau), levels of phosphorylated tau (P-tau) isoforms (including tau phosphorylated at 181 (P-tau181), 217 (P-tau217), and 231 (P-tau231)), glial fibrillary acidic protein (GFAP), and neurofilament light chain (NfL)). For example, it has been found that subjects treated with elenbecestat (E2609), a β -site amyloid precursor protein cleaving enzyme (BACE) inhibitor, who had amyloid baseline positron emission tomography (PET) standard uptake value ratios (SUVr values) of 1.4 to 1.9, exhibited the greatest slowing of cognitive decline while on treatment. See Lynch, S. Y. et al. "Elenbecestat, a BACE inhibitor: results from a Phase 2 study in subjects with mild cognitive impairment and mild-to-moderate dementia due to Alzheimer's disease." Poster P4-389, Alzheimer's Association International Conference, July 22-26, 2018, Chicago, IL, USA. Similarly, it has been found that subjects having a baseline florbetapir amyloid PET SUVr levels below 1.2 do not exhibit enough cognitive decline to be detectable, whereas subjects having SUVr levels above 1.6 appear to correlate with a plateau effect in which amyloid level has reached a saturation level and treatment does not result in a change of cognitive measures. See Dhadda, S. et al., "Baseline florbetapir amyloid PET standard update value ratio (SUVr) can predict clinical progression in prodromal Alzheimer's disease (pAD)." Poster P4-291, Alzheimer's Association International Conference, July 22-26, 2018, Chicago, IL, USA.

[0160] In some embodiments, a subject (or patient) described as having early AD may exhibit AD severity ranging from mild cognitive impairment due to AD – intermediate likelihood to mild Alzheimer’s disease dementia. Subjects with early AD include subjects with mild Alzheimer’s disease dementia as defined herein and subjects with mild cognitive impairment (MCI) due to AD – intermediate likelihood as defined herein. In some embodiments, subjects with early AD have MMSE scores of 22 to 30 and Clinical Dementia Rating (CDR) global range 0.5 to 1.0. Other methods for detecting early AD disease may employ the tests and assays specified below, including the National Institute of Aging-Alzheimer’s Association (NIA-AA) core clinical criteria for probable Alzheimer’s disease dementia in McKhann, G.M. *et al.*, “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging – Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease.” *Alzheimer Dement.* 2011; 7:263-9. Other methods include CDR-SB, ADCOMS, the MMSE, ADAS-Cog, ADAS MCI-ADL, modified iADRS, WMS-IV LMI, WMS-IV LMII, or the like. In some embodiments, a subject with early AD has evidence of elevated amyloid in the brain or a positive amyloid load. In some embodiments, elevated amyloid in the brain or a positive amyloid load is indicated and/or confirmed by PET assessment. In some embodiments, elevated amyloid in the brain or a positive amyloid load is indicated and/or confirmed by a CSF assessment of markers such as A β 1-42 (e.g., a soluble CSF biomarker analysis). In some embodiments, elevated amyloid in the brain or a positive amyloid load is indicated and/or confirmed by measuring the level of p-tau181. In some embodiments, elevated amyloid in the brain or a positive amyloid load is indicated and/or confirmed by an MRI. In some embodiments, elevated amyloid in the brain or a positive amyloid load is indicated by retinal amyloid accumulation. In some embodiments, more than one assessment method is used.

[0161] In some embodiments, a control subject (or patient), untreated AD subject (or patient), or an untreated control subject (or patient) is a subject (or patient) that is not being treated or has been treated for Alzheimer's disease. In some embodiments, a control subject has Alzheimer's disease. In some embodiments, the control subject has early Alzheimer's disease, or pre-Alzheimer's disease. In some embodiments, the control subject has Alzheimer's disease and is not treated with an anti-A β protofibril antibody.

[0162] As may be appreciated, a trained predictive model consistent with disclosed embodiments can be used to screen or select patients for inclusion in a clinical trial. Brain tau or amyloid β status data can be predicted for a candidate patient using baseline data acquired for that candidate patient. The candidate patient can be included in the study when the predicted brain tau or amyloid β status satisfies a selection criterion. In various embodiments, the selection criterion can depend on a predicted brain tau or amyloid β level, satisfaction of certain demographic factors, a biomarker measurement, or another suitable measure. For example, a patient may be included in a clinical trial when a predicted brain tau or amyloid β level exceeds a threshold value or is within a specified range. Such patients may have a greater need for treatment (as they might otherwise experience greater cognitive decline).

[0163] As may be appreciated, a trained predictive model consistent with disclosed embodiments can be used in clinical trial design. As described herein, a clinical trial population can be enriched with patients likely to benefit from a treatment. In particular, a trained predictive model can be used to screen or select patients for inclusion in a clinical trial. The patients selected can be those likely to have at least a minimum threshold amount of brain tau or amyloid β deposits, or at most a maximum amount of brain tau or amyloid β deposits. Such patients may be suitable targets for treatments adapted to removing or mitigating tau (or amyloid β) deposits. As may be appreciated, treatment effect, study size, and study power can be related. By selecting patients likely to be suitable treatment targets,

fewer patients can be enrolled, or study power can be increased, or detectable treatment effect size reduced, or some combination of the foregoing.

[0164] Furthermore, the benefits of treatment for such patients may be more readily apparent than for less-afflicted patients. Because the effects of treatment are more apparent (e.g., treatment effects are larger), screening or selecting patients using a trained predictive model can enable an improved clinical trial design: the number of patients enrolled can be reduced, the minimum detectable treatment effect can be increased, study power can be increased, or some combination of the foregoing.

[0165] As may be appreciated, predicted brain tau or amyloid β status can be used to evaluate the effect of a treatment for neurological disease, dysfunction, or injury in a clinical trial or real-world clinical settings. The clinical trial may include multiple participants. The participants can be screened for satisfaction of demographic factor or biomarker data, and baseline data can be acquired for each participant. The participants can be assigned to either a control or a treatment group of the study.

[0166] Using the trained predictive model, brain tau or amyloid β status can be predicted for one or more participants in the treatment group of the clinical trial and used as a covariate in analyzing the results of the clinical trial.

[0167] For example, the clinical trial can concern an Alzheimer's treatment. The trained predictive model can be used to predict brain tau or amyloid β status for at least some patients assigned to the treatment group of the clinical trial. The predicted brain tau or amyloid β status can be used as a covariate in determining an effect of the Alzheimer's treatment.

[0168] **FIG. 3** depicts a process 300 for treating brain tau or amyloid β deposits, consistent with embodiments of the present disclosure. In some embodiments, process 300 is described as being performed using platform 100. However, process 300 can also be performed at least in part using another computing system. Process 300 can be used to predict brain tau or brain

amyloid β status. When predicting brain amyloid β status, medical record data for a patient may include, or be derived from, A β -SUVR levels or centiloid levels. In some such embodiments, the predictive models may output predicted A β -SUVR levels or centiloid levels, or A β categories (e.g., A β +, A β -, or the like). When predicting brain tau status, medical record data for a patient may include, or be derived from, tau-PET SUVR levels. In some such embodiments, the predictive models may output predicted tau-SUVR levels, or tau categories (e.g., tau positive, tau negative, or the like). Examples of the development of predictive brain tau models and brain amyloid β models are provided herein.

[0169] Process 300 may involve a step 310 of obtaining patient medical data (e.g., subject data for a patient). In some embodiments, the subject data can be obtained from medical record data of the patient. Components of platform 100 (e.g., ETL engine 110, dataset creation engine 115, or the like) can obtain the subject data, consistent with disclosed embodiments. In some embodiments, the subject data may include biomarker data, as described herein. In some embodiments, the subject data may include biomarker data and at least one of cognitive measurement data, genomic data, imaging data, demographic data, or the like.

[0170] Process 300 may involve a step 320 of predicting brain tau or amyloid β status. Step 320 may involve predicting brain tau or amyloid β status for the patient by applying at least a portion of the obtained subject data (e.g., obtained in step 310) to a predictive model. In some embodiments, the predictive model may be trained to predict brain tau or amyloid β status using a training dataset. The training dataset can include training examples. A training example can include subject data and label information, as described herein. In some embodiments, the subject data used to train the predictive model can include the same type of information as the subject data obtained in step 310. For example, when the predictive model was trained using subject data including pTau217R values and demographic information,

subject data obtained in step 310 can include pTau217R values and the same demographic information.

[0171] In some embodiments, at least some of the training subjects may satisfy an impairment criterion, as described herein. It will be appreciated that by predicting levels of brain tau or amyloid β deposits for patients that satisfy an impairment criterion (e.g., and therefore more likely to have elevated levels of brain tau or amyloid β deposits) the accuracy of such predictions may be improved.

[0172] In some embodiments, step 320 may be executed as described regarding process 200. Additionally, or alternatively, step 320 may be executed with other suitable predictive models. As described herein, the predicted brain tau or amyloid β status (e.g., predicted in step 320) may indicate an elevated level of brain tau or amyloid β deposits. The prediction can concern the brain as a whole, brain regions, collections of brain regions, or other portions of the brain. Such portions can be associated with certain Braak stages or major cortical lobes of the brain (e.g., cortical, parietal, occipital, lateral temporal, medial temporal) and entire cortical grey matter (e.g., whole cortical grey matter or WCGM). Thus, the predicted brain tau or amyloid β status may enable treatment and/or management of neurological disease, disorder, or injury. For example, the predicted brain tau or amyloid β status may include an indication that the patient seeks further testing (e.g., cognitive assessment testing, additional biomarker testing, imaging) or monitoring (e.g., follow-up with a clinician). In some embodiments, the predicted brain tau or amyloid β status may include data for determining or predicting a progression of neurological disease, disorder, or injury for the patient.

[0173] In some embodiments, process 300 may proceed to step 330 based on the predicted brain tau or amyloid β status in step 320. For example, the predicted brain tau or amyloid β status may indicate that a patient has an elevated level of tau deposits in one or more brain regions. In response, process 300 may proceed to step 330. Otherwise (e.g., should the patient

be predicted to have a low level of brain tau or amyloid β deposits), process 300 can terminate.

[0174] Process 300 may involve an optional step 330 of confirming predicted brain tau or amyloid β status. For example, the predicted brain tau or amyloid β status can be confirmed using suitable imaging techniques (e.g., a PET scan using a suitable tau radiotracer, or another suitable confirmatory method). As may be appreciated, performing such imaging in response to a prediction of elevated levels of tau deposits may enable more efficient use of scarce clinical resources.

[0175] Process 300 may involve an optional step 340 of providing treatment. Step 340 may be based on the predicted brain tau or amyloid β status (e.g., step 320). For example, based on the predicted brain tau or amyloid β status, treatment(s) may be provided to the patient, or instructions or recommendations for such treatments can be provided by platform 100 (e.g., to a user of platform 100). In some embodiments, treatment may be provided based on other factors in addition to the brain tau or amyloid β status, such as factors in the patient's medical history.

[0176] In some embodiments, providing treatment may involve administering a therapeutic agent to the patient. For example, the predicted brain tau or amyloid β status for a patient may contribute to determining a therapeutic agent to administer to the patient. It will be appreciated that differing therapeutic agents may be administered to patients based on the predicted brain tau or amyloid β status. A therapeutic agent may refer to any substance, treatment, or compound capable of treating, curing, mitigating, or preventing diseases or conditions. For example, a therapeutic agent may be a biologically active compound capable of treating dementia-related conditions. In some embodiments, the therapeutic agent may be an anti-amyloid beta protofibril antibody. For example, the antibody may be lecanemab. Sequence and other information on lecanemab can be found in PCT/US2022/073576, which

is incorporated by reference in its entirety. Exemplary dosing regimens for lecanemab are disclosed in International Application No. PCT/US2024/033125, which is also incorporated herein by reference in its entirety. As another example, the therapeutic agent may be an anti-tau antibody such as E2814. Sequence information and exemplary dosing for E2814, including as a combination therapy with lecanemab, are disclosed in International Application Nos. PCT/IB2021/000937 and PCT/US2022/079509, each of which are incorporated by reference herein in their entireties.

[0177] In some embodiments, providing treatment may involve for monitoring a treatment effect over time. For example, such monitoring can include determining whether a treatment (e.g., Lecanemab) reduces the predicted brain tau or amyloid β levels over time, or reduces a rate of increase of predicted brain tau or amyloid β levels over time, or the like.

[0178] In some embodiments, steps of process 300 can be repeated periodically. For example, step 310 and 320 can be predicted periodically to monitoring the brain tau or amyloid β status of the patient. For example, based on brain tau or amyloid β status, a treatment may not initially be indicated for a patient. The patient may be monitored by repeating the steps of process 300. As the condition of the patient progresses, that progression may be indicated in the patient's predicted brain tau or amyloid β status. Once the patient satisfies a treatment condition, treatment may be provided to the patient. Similarly, once treatment is being provided to a patient, steps of process 300 can be repeated periodically to monitor the effectiveness of the treatment (e.g., a decrease in a degree of neurofibrillary tangles, or another indication described herein). As may be appreciated, the disclosed embodiments can permit such a monitoring strategy, which may not be feasible with existing assessment tools, such as tau-PET imaging.

[0179] It will be appreciated that the disclosed embodiments provide improvements to predicting brain tau or amyloid β status and providing treatment (including improved

selection for patients likely to respond to treatment) for elevated levels of brain tau or amyloid β deposits. As described herein, the disclosed embodiments may be provided as a screening tool (e.g., for clinicians or the like) to improve their ability to identify individuals having elevated levels of brain tau or amyloid β deposits, thereby providing an entry point for triage. The disclosed embodiments may assist in early identification of neurological disease, disorder, or injury, and may alert clinicians to the possibility of undetected risks in patients. The disclosed embodiments may also assist in the design and execution of clinical trials of new compounds intended to combat the deposition of tau (or A β) in the brain of patients. The disclosed embodiments may also provide improvements in monitoring and managing such neurological disease, disorder, or injury by predicting data that may indicate a patient should seek follow up (e.g., with a specialist) and/or seek further testing.

[0180] Examples

[0181] Multiple investigations were performed into the training and use of predictive models consistent with disclosed embodiments. These investigations concerned prediction of brain tau or amyloid β status using subject data, as described herein.

[0182] A first study concerned generating predictive models for detecting tau neurofibrillary tangles along the Braak stages. Neurofibrillary tangles can be a pathological hallmark of Alzheimer's Disease, and the disclosed embodiments may assist in developing non-invasive methods for detecting tau tangles (e.g., to enable efficient screening for patients in AD clinical trials). The first investigation involved generating predictive models for detecting tau tangles along the Braak stages using demographics, ApoE4 status, and one or more of cognitive function assessments, plasma pTau181, and brain regional MRI measures (e.g., volume, area and cortical thickness).

[0183] **FIG. 4** illustrates a diagram 400 of the progression of Alzheimer's Disease along the Braak stages, consistent with disclosed embodiments. Diagram 400 indicates various regions of the brains encompassed by Braak stage 1, stage 2, stage 3, stage 4, stage 5, and stage 6.

[0184] The data in the first investigation included baseline regional MK6240 tau-PET SUVR data for amyloid-positive patients from a clinical trial. The tau positivity threshold for each Braak stage was defined via the one-sided upper 99% confidence limit of a subset of 70 subjects with amyloid-PET Centiloid < 30 (representing negative control). Cognitive function assessments included ADAS-COG-14, CDR-SB, MMSE, and their sub-scores. Plasma pTau 181 was measured using the Simoa[®] assay. Subjects received a 3.0 Tesla (T) structural MRI at baseline. Brain MRI data (volume, area, and cortical thickness) were generated for various brain regions of interest using the Desikan-Killiany atlas (Desikan et al., 2006), resulting in 207 regional measures. Cortical thickness values are represented in millimeters (mm). The volume (mm³) and area (mm²) were normalized (divided) by the intra-cranial volume to reduce inter-subject variability.

[0185] In the first investigation, to reduce the dimensionality of the data, structural brain network (SBN) modules, and hub regions were derived using data from other clinical trials via the MEGENA algorithm. Such SBN modules and hub regions may represent clusters and regions with the strongest spatial correlation to neighboring regions. **FIGs. 5A to 5C** illustrate exemplary SBN hubs and modules, consistent with disclosed embodiments. **FIG. 5A** depicts a module representing the inferior parietal cortical thickness, and **FIG. 5B** depicts a module representing precentral cortical thickness. **FIG. 5C** depicts brain hubs that are among the top predictors for discriminating Braak 3-4 and 5-6.

[0186] To detect tau positive in Braak stages 3-6, signatures were derived via stochastic gradient boosting (e.g., an ensemble tree-based machine learning algorithm). To further discriminate tau positivity in Braak stages 3-4 and 5-6, signatures were derived via Bayesian

ordinal logistic regression with a Student's t prior. Features for deriving the signatures came from data types including clinical cognitive function assessments, plasma pTau181, MRI-based SBNs and hubs, amyloid PET centiloid, demographic data (e.g., age, gender, BMI), and genomic data (e.g., ApoE4 status). Data was randomly split into a 70-30 training-testing split. Prediction performance was assessed via 10-fold cross-validation in the training set followed by evaluation in the test set.

[0187] **FIG. 6** illustrates a table summarizing patient data, consistent with disclosed embodiments. **FIG. 6** indicates tau positivity status in various Braak stages for subjects according to clinical diagnosis (e.g., Mild Cognitive Impairment or mild Alzheimer's Disease), gender (e.g., female or male), ApoE4 status (e.g., Non-E4 or E4 carrier), age, BMI, and MMSE score.

[0188] **FIG. 7** illustrates the distribution of SUVR values for subjects with different tau positivity status (e.g., tau positivity in Braak stage 0, Braak 1-2, Braak 3-4, or Braak 5-6), consistent with disclosed embodiments.

[0189] **FIG. 8** illustrates a table 800 displaying the performance of tau positivity prediction, consistent with disclosed embodiments. Signatures comprising clinical cognitive assessments, ApoE4 status, and demographics achieved 76% accuracy for detecting tau positivity in Braak 3-6, and 60.8% accuracy for discriminating Braak 3-4 versus 5-6. Adding plasma pTau181 improved the accuracy to 82.2% ($p<0.05$) in Braak 3-6, and to 67.6% accuracy ($p<0.05$) for Braak 3-4 versus 5-6. Detection accuracy of tau positivity in Braak 3-4 versus the spread to Braak 5-6 improved from 66.8% to 74.6% ($p<0.05$) when combining MRI data with plasma pTau181.

[0190] **FIGs. 9A-9C** illustrate top predictors for detecting tau positive subjects in Braak 3-6 using a stochastic gradient boosting machine model, consistent with disclosed embodiments. **FIG. 9A** illustrates the relative influence of features using clinical (e.g., cognitive function)

data and plasma pTau181 data. **FIG. 9B** illustrates the relative influence of features using plasma pTau181 and MRI data. **FIG. 9C** illustrates the relative influence of features using clinical (e.g., cognitive function), plasma pTau181, and MRI data. In **FIGs. 9A-9C**, HV.L refers to hippocampal volume (left), VCIPCR refers to inferior parietal cortical thickness (right), VVEL refers to entorhinal cortex, VCIPCL refers to inferior parietal cortical thickness (left), VCPRCR refers to precentral cortical thickness (right), ADCDRL refers to delayed word recall, and ADCRL refers to word recall.

[0191] **FIGs. 10A-10C** illustrate individual conditional expectation profiles of some features predictive of tau positivity, consistent with disclosed embodiments. **FIGs. 10A-10C** illustrate individual conditional expectation (ICE) profiles for each subject (in grey) and the average subject (in black). The vertical axis represents the normalized probability of tau positivity, after normalizing for the probability at the minimum value of the feature. Inter-subject heterogeneity in the ICE profiles may be due to the strong interaction between the predictors. The stochastic gradient boosting algorithm can account for these nonlinear relationships and interactions without prior assumptions.

[0192] **FIGs. 11A-11C** illustrate heat maps displaying interactions between some features predictive of tau positivity (e.g., with heat map intensity corresponding to the probability of tau positivity), consistent with disclosed embodiments. **FIG. 11A** illustrates a heat map displaying interactions between plasma pTau181 and delayed word recall. **FIG. 11B** illustrates a heat map displaying interactions between plasma pTau181 and MRI inferior parietal cortical thickness (right). **FIG. 11C** illustrates a heat map displaying interactions between MRI inferior parietal cortical thickness (right) and delayed word recall. It will be appreciated that the interaction prediction profiles demonstrate the strong dependence between some key features for predicting tau positivity. For example, among patients that are relatively weak in delayed word recall, those with high plasma pTau181 are more likely to be

tau-positive. Similarly, patients with higher plasma pTau181 are more likely to be tau-positive if they have less thickness in the inferior parietal cortex.

[0193] **FIGs. 12A-12B** illustrate plasma pTau181 as a function of tau positivity status, consistent with disclosed embodiments. **FIG. 12A** illustrates that plasma pTau181 monotonically increases with tau presence along Braak stages (e.g., Braak 0-1-2, Braak 3-4, and Braak 5-6). **FIG. 12B** illustrates plasma pTau181 as a function of tau positivity status, grouped by tertiles of inferior parietal cortical thickness. **FIG. 12B** illustrates that when combined with the inferior parietal cortical thickness, plasma pTau181 levels are more strongly differentiated between tau presence in Braak 3-4 versus Braak 5-6 for most patients, with the accuracy increasing from 67.6% to 74.6% ($p < 0.05$).

FIGs. 13A-13C illustrate predictors for differentiating tau positive subjects in Braak stages 3-4 and Braak stages 5-6 using Bayesian ordinal logistic models, consistent with disclosed embodiments. **FIG. 13A** illustrates predictors using clinical data and plasma pTau181 data. **FIG. 13B** illustrates predictors using plasma pTau181 data and MRI data. **FIG. 13C** illustrates predictors using clinical data, plasma pTau181 data, and MRI data. In **FIGs. 13A-13C**, VCIPCR refers to inferior parietal cortical thickness (right), VCPRCR refers to precentral cortical thickness (right), VVPCCL refers to posterior cingulate cortical volume (left), ADCDRL refers to delayed word recall, ADCCMD refers to commands, ADCOF refers to naming objects, ADCCP refers to constructional praxis, and CDR0101 refers to memory.

[0194] **FIGs. 14A-14E** illustrate top MRI predictors. In **FIGs. 14A-14E**, VCIPCR refers to inferior parietal cortical thickness (right), and VCPRCR refers to precentral cortical thickness (right). **FIG. 14A** illustrates odds ratios for predicted tau positivity using plasma pTau181, ApoE4 count, and MRI data (e.g., VCIPCR, VCPRCR). **FIG. 14B** illustrates VCIPCR as a function of tau positivity status in Braak stages 0-2, stages 3-4, and stages 5-6. It will be

recognized that tau positive subjects in Braak stages 3-4 & Braak stages 5-6 had

progressively less thickness in the inferior parietal cortex, as reflected by the odds ratio < 1 .

FIG. 14C illustrates VCPRCR as a function of tau positivity status in Braak stages 0-2, stages 3-4, and stages 5-6. **FIG. 14D** illustrates VCPRCR as a function of tau positivity status in Braak stages 0-2, 3-4, and 5-6, grouped by VCIPCR tertiles. It will be appreciated that pre-central cortical (PCC) thickness on its own did not differentiate the tau positive Braak stages. When stratified by inferior parietal cortical (IPC) thickness, there were increasing levels of PCC thickness in tau positive Braak stages 3-6. Such a trend may have resulted from increased levels of inflammation/gliosis due to tau reaching Braak stages 5-6 in the PCC. That is, atrophy in the IPC due to tau positivity in Braak 3-4 may precede brain inflammation. Tau presence in Braak 5-6 in the inflamed brain may then manifest in the form of greater thickness in Braak 6 regions such as the PCC, cuneus, and postcentral cortex. The first cluster of subjects represented by an IPC thickness of 1.724-2.217 may indicate a more advanced disease stage, with more atrophy in the parietal regions and more cases with tau in Braak stages 5-6. The third cluster of subjects represented by an IPC thickness of 2.376-2.691 may indicate a less advanced disease stage, with less atrophy in the parietal regions and more cases with tau positivity in Braak stages 0-2. **FIG. 14E** illustrates subjects represented by an IPC thickness of 2.217-2.376 (e.g., the second cluster of subjects as illustrated in **FIG. 14D**). In **FIGs. 14D-14E**, Braak stages 5-6 may be thicker than Braak 0-2 or Braak 3-4 within a given cluster of inferior parietal cortical thicknesses, which may be due to tau-related inflammation/gliosis. Additionally, within a given cluster of inferior parietal cortical thicknesses, Braak stages 0-2 may be thinner as neurodegeneration may have already occurred or may be occurring.

[0195] **FIGs. 15A-15B** show similar trends in the other Braak 6 regions. **FIG. 15A** illustrates postcentral cortical thickness as a function of tau positivity status in Braak stages 0-2, Braak

3-4, and Braak 5-6, grouped by VCIPCR tertiles, consistent with disclosed embodiments.

FIG. 15B illustrates cuneus cortical thickness as a function of tau positivity status in Braak stages 0-2, Braak 3-4, and Braak 5-6, grouped by VCIPCR tertiles, consistent with disclosed embodiments.

[0196] **FIG. 16** illustrates a table displaying how, in the first study, adding Amyloid PET centiloid data improved tau positivity prediction, consistent with disclosed embodiments. In some embodiments, adding Amyloid PET data to plasma pTau181 and/or clinical assessments and/or MRI can yield more accurate predictions of tau positivity.

[0197] **FIGS. 17A-17B** illustrate how amyloid PET levels can play a complementary role for predicting tau positivity. **FIG. 17A** illustrates Amyloid PET centiloid levels as a function of tau positivity status in Braak stages 0-2, stages 3-4, and stages 5-6, grouped by plasma pTau181 levels. In this study, Braak stages 3-4 and Braak 5-6 tau positive subjects with low pTau181 had high Amyloid levels. Similarly, tau-negative subjects with high pTau181 had low Amyloid levels. **FIG. 17B** illustrates an interaction profile demonstrating how tau positivity status of subjects with high pTau181 can depend greatly on their Amyloid levels.

[0198] Accordingly, the first investigation demonstrated detection of tau deposition throughout the brain along the Braak stages using non-invasive measures such as plasma pTau181, cognitive performance, and brain structural MRI data. These signatures can be used for patient screening in clinical trials. In some cases, these signatures may be followed by confirmations with other assessments, such as imaging (e.g., tau-PET) or the like, in a subset of patients. These signatures can also be used in prior clinical studies that did not include tau-PET assessments for exploratory evaluation of the association of tau with clinical progression, subgroup effects, or the like. Further, it will be appreciated that, in some cases, adding Amyloid-PET data can yield better predictions of tau positivity.

[0199] A second study investigated the effectiveness of plasma pTau217R in predicting continuously valued measurements of regional tau levels (e.g., tau-PET SUVR measurements) and identifying subjects with different levels of tau accumulation. The study further sought to differentiate individuals with varying levels of tau accumulation within a cohort diagnosed with amyloid- β positive ($A\beta+$) mild cognitive impairment (MCI) or mild Alzheimer's disease (referred to collectively as early Alzheimer's disease). Plasma pTau217 and non-phosphorylated tau217 concentrations were quantified via immunoprecipitation-mass spectrometry. Predictive models for MK6240 tau-PET SUVR were developed and validated using a 60-40 random split of a clinical trial cohort comprising 242 amyloid- β positive early Alzheimer's disease individuals. A stochastic gradient boosting algorithm was employed to construct models for concurrently predicting SUVR values across various brain regions. Additional analyses explored whether the integration of additional predictors (e.g., cognitive assessments, fluid biomarkers, structural MRI, and amyloid PET) into the model could improve the prediction performance of pTau217R. Model performance was cross-validated within the training set and evaluated further in a hold-out test set.

[0200] In this second study, pTau217R-based models predicted tau-PET uptake across various brain regions, with R^2 values ranging from 0.49 to 0.65. The maximum SUVR values predicted across these brain regions fell within the range of 1.87 to 2.3. The area under the receiver operating characteristic curve (AUROC) for detecting tau presence ranged from 84% to 95% across the six Braak stages and cortical regions, maintaining performance at higher tau accumulation levels. In this study, integrating additional predictors did not improve pTau217R's performance. Using pTau217R alone in predicting tau-PET SUVR reduced the need for tau-PET scans by up to 65%, particularly in identifying low tau concentrations within cortical grey matter, while maintaining a 5% false negative rate. As may be appreciated, the techniques applied in this study could enable pathological disease staging

(i.e., tau load) using measurements of plasma pTau217R and potentially reduce the need for tau-PET scans.

[0201] The data used for constructing and validating the prediction models included A β + early Alzheimer's disease subjects with an objective cognitive memory impairment and a wide range of tau-PET SUVR values in different brain regions. The training and validation data sets for constructing the prediction models were derived from the 60-40 random split of 242 early Alzheimer's disease subjects from the Clarity AD clinical study, for whom both tau-PET and plasma biomarker data were available (A Study to Confirm Safety and Efficacy of Lecanemab in Participants With Early Alzheimer's Disease; NCT03887455).

[0202] **FIG. 18A** provides a summary of key demographic and clinical characteristics of the subjects included in the data used for constructing and validating the prediction models, consistent with disclosed embodiments. As apparent in **FIG. 18A**, the distribution of tau-PET SUVR values in the WCGM and MTL spanned a broad spectrum, ranging from low to intermediate to high tau levels. Except for age, none of the characteristics showed significant differences between the training and test sets. In **FIG. 18A**, SD refers to standard deviation.

[0203] All participants underwent a structural MRI scan. Brain MRI data, including volume, area, and cortical thickness across various brain regions of interest, were derived using the Desikan-Killiany atlas, resulting in 207 regional measures. An imaging pipeline was employed for image processing. Cortical thickness values were denoted in millimeters (mm), while volume (cubic millimeters) and area (square millimeters) were normalized by intracranial volume to mitigate intersubject variability and adjust for differences in head size across regions. To streamline the analysis and focus on pertinent data, prediction models utilizing the MRI data were constructed specifically targeting the 18 structural brain network (SBN) modules and 45 hub regional measures identified in U.S. Provisional Patent

Application No. 63/561,285, and incorporated herein by reference, thereby reducing redundancy and dimensionality within the dataset.

[0204] Tau-PET SUVR values from brain regions defined using the Hammers atlas in PMOD were combined into volume-weighted average composite cortical regions, including frontal, occipital, parietal, lateral temporal, and medial temporal regions, as well as the six Braak stage regions and WCGM. SUVR values from amyloid PET scans (89% florbetaben, 10% florbetapir, and 1% flutemetamol) were standardized to Centiloids for consistency and comparability.

[0205] In this study, plasma levels of A β 42, A β 40, pTau217, and nonphosphorylated (np) Tau217 were measured using a high-throughput, mass spectrometry-based analytical platform. The ratios A β 42/A β 40 and pTau217/npTau217 were calculated. Blood samples were collected into K2EDTA tubes, processed into plasma, and stored at -80°C until analysis. Plasma aliquots were shipped on dry ice to a commercial laboratory for processing and analysis. The pTau217/npTau217 ratio (pTau217R) was used for subsequent analyses to normalize inter-individual differences.

[0206] Plasma levels of pTau181, neurofilament light chain (NfL), and glial fibrillary acidic protein (GFAP) were assessed using commercially available single molecule array (Simoa®) assay kits: the Simoa® pTau-181 Advantage V2 assay kit, the Simoa® NF-Light assay kit, and the Simoa® GFAP discovery kit, respectively.

[0207] In this study, clinical assessments encompassing both cognitive and functional domains were also evaluated to ascertain their viability as predictors in the prediction models incorporating pTau217R. These assessments included the MMSE, ADAS-Cog-14, CDR-SB, ADCS-ADL, and their respective sub-scores.

[0208] A model for predicting regional tau-PET SUVR for each patient was initially constructed from the training set. The predictive model was a Stochastic Gradient Boosting

(SGB) model and used plasma pTau217R, age, sex, and tau brain region as input subject data. The predictive model output predictions for select cortical regions (frontal, parietal, occipital, lateral temporal, and medial temporal), along with regions delineated by the six Braak stages. Additional predictive models were developed that accepted additional subject data, including clinical assessments, volumetric MRI (vMRI) measures, amyloid PET Centiloid, and other plasma biomarkers (pTau181, A β 42/A β 40, GFAP, and NfL).

[0209] To mitigate overfitting, the model underwent internal refinement through hold-out datasets and cross-validation techniques. In this study, up to 1000 decision trees were assembled with up to three-way interactions among predictors. To ascertain the significance of individual predictors within the prognostic framework, the impact of such individual predictors on reducing mean squared error (MSE) when employed as root nodes for tree splitting in the SGB algorithm was assessed. These measures were then normalized, yielding predictor rankings and relative influences scaled uniformly from 0 to 100%.

[0210] In this study, the prediction performance of the models was evaluated. A first evaluation was conducted through 10 iterations of 10-fold cross-validation within the training set. A second evaluation was conducted using the validation set. This evaluation entailed measuring key measures, including the coefficient of determination (R^2), MSE, and MAE for observed versus predicted SUVR values across each brain region. The correlation of observed versus predicted SUVR values was compared between the models via Hittner's two-sided test for dependent correlations. Additionally, the potential added benefit of incorporating ApoE4 allelic count in the models and the use of pTau217 concentration instead of pTau217R were assessed within this framework.

[0211] To estimate the upper limit of reliable prediction for tau-PET SUVR across different brain regions, a Richards 5-parameter generalized logistic model was fitted using data from subjects within the validation set. This model encompasses five key parameters: the lower

and upper asymptotes, the slope representing the steepness of the curve, the inflection point marking where the curvature transitions, and the degree of asymmetry indicating the shape of the sigmoidal curve. The bottom asymptote was set at 1, reflecting the expected background tau level. The upper asymptote parameter was estimated to delineate the maximum SUVR prediction limit.

[0212] To evaluate the versatility of the tau-PET SUVR prediction models across varying rates of tau accumulation along the spectrum of early Alzheimer's disease, the ability of the models to predict tau status (i.e., whether the predicted SUVR value falls below or above a threshold) was evaluated for each of the cortical and Braak stage regions for two SUVR thresholds: the threshold for tau positivity, which varied for each region, and a higher threshold of 1.5. This evaluation was conducted within the validation set, employing receiver operating characteristic (ROC) curves and relevant metrics such as the area under the ROC curve (AUROC), sensitivity, and specificity. This performance evaluation was then repeated for each demographic subgroup (i.e., age, sex, and ApoE4 status) in the validation set to determine whether the models performed consistently across all the subgroups. For simplicity, only WCGM and MTL regions were considered for this assessment.

[0213] Determining the SUVR threshold for tau positivity in each region involved analyzing the data from a subset of 50 subjects within the training cohort. These subjects had an amyloid PET Centiloid below 15, indicative of minimal to no amyloid accumulation. This subset was chosen to represent the normative tau population with background (no/low) tau levels. The threshold for tau positivity was established as the upper 95% confidence limit, calculated using the formula: $\text{median} + 1.645 \times 1.4826 \times \text{MAD}$. Here, 1.645 represents the upper 5% limit of the standard normal distribution, MAD signifies the median absolute deviation, and $1.4826 \times \text{MAD}$ provides a suitable approximation to the standard deviation (SD). Given the left-skewed distribution of SUVRs, which deviates from normality, the

utilization of median and MAD, resilient to outliers and distributional shape, enables the application of the normal approximation formula for determining the tau positivity threshold.

[0214] To assess the potential savings in tau-PET scans by employing blood-based biomarkers through the SUVR prediction models to identify subjects with tau accumulation below a specific threshold, the sensitivities of 95%, 90%, and 80% were chosen, allowing for 5%, 10%, and 20% false negative rates, respectively. The specificity of these models was then determined at the tau positivity threshold and a higher threshold of 1.5 SUVR. In a scenario where tau-PET scans are conducted only on subjects predicted to have tau levels above a specific threshold of interest, the reduction in PET scans reflects the accurate identification of subjects that have tau levels below that threshold. This alignment is based on the specificity of the model, which corresponds to 1 minus the false positive rate. For simplicity, these evaluations were carried out for only the WCGM and MTL regions.

[0215] The training and test sets included 144 and 98 A β + subjects, respectively. Among these, approximately two-thirds were MCI due to AD, while the rest had mild Alzheimer's disease. The distribution of tau-PET SUVR values across the WCGM and MTL was not significantly different between the training and test sets, as anticipated, given that these were random splits from the same cohort. The interquartile range (25th to 75th percentiles) of tau-PET SUVR distribution in WCGM ranged from 0.98 to 1.55 for the training set, and from 0.96 to 1.76 for the validation set. For MTL, the interquartile range was 1.04 to 1.97 for the training set and 1.05 to 1.98 for the validation set. **FIG. 18B** depicts nonlinear patterns in the relationship between plasma pTau217R and tau-PET SUVR values across select cortical regions (lateral temporal 1802, parietal 1804, WCGM, 1806, occipital 1808, MTL 1810, frontal 1812), consistent with disclosed embodiments. **FIG. 18C** depicts similar nonlinear patterns in the relationship between plasma pTau217R and tau-PET SUVR values in the six

Braak stage regions (Braak IV 1814, Braak I 1816, Braak V 1818, Braak III 1820, Braak II 1822, Braak VI 1824), consistent with disclosed embodiments.

[0216] Key demographic and clinical characteristics, including diagnosis, age, sex, race, ApoE4 status, and MMSE scores, are summarized in **FIG. 18A**. Since the training and validation sets were derived from a randomized 60-40 split of a clinical study, no significant differences were observed in any variables between the two sets, except for age. Wilcoxon test was used for comparing age, MMSE, and tau-PET SUVR. The chi-squared test was used for comparing diagnosis, sex, race, and ApoE4 status.

[0217] **FIG. 19** depicts an overview of the prediction performance of SUVR models across various brain regions through cross-validation, consistent with disclosed embodiments. In particular, **FIG. 19** shows R^2 values of the SUVR models across cortical regions and Braak stages, as derived via 10 iterations of 10-fold cross-validation. Notably, the model exclusively leveraging plasma pTau217R, in conjunction with age, sex, and tau brain region, showed superior performance compared to models using alternative predictors such as clinical assessments, MRI measures, amyloid PET Centiloid values, and other plasma biomarkers ($p < 0.05$, as per Hittner's test on correlations). This model achieved robust R^2 values ranging from 0.338 to 0.635 for cortical regions and 0.236 to 0.639 for the six Braak stage regions. Furthermore, integration of these additional predictor types (e.g., clinical assessments, MRI measures, amyloid PET Centiloid) into the pTau217R-based model did not yield improvements in the R^2 values, a trend consistent across other performance metrics, including root mean squared error (RMSE) and MAE.

[0218] As depicted in **FIGs. 20A** through **20E**, pTau217R appeared as the dominant predictor in all the multivariate models, consistent with disclosed embodiments. The relative influence of predictors is depicted for models incorporating pTau217R alone (**FIG. 20A**), pTau217R alongside clinical assessments (**FIG. 20B**), structural MRI measures (**FIG. 20C**),

amyloid PET Centiloid (**FIG. 20D**), and other plasma biomarkers (**FIG. 20E**), respectively.

As shown, plasma pTau217R emerged as the predominant predictor, exerting at least 50% relative influence across all scenarios. Following closely was the tau brain region, contributing approximately 20% relative influence, reflecting the disparities in the tau-PET SUVR distribution across different regions.

[0219] In these figures, CDR-SB is the clinical dementia rating sum of boxes score; ADCSADL is the AD cooperative study activities of daily living score; ADCDR is the delayed word recall score; ADAS.14 is the Alzheimer's Disease Assessment Scale-Cognitive Subscale 14 score; MMRL is the recall score; CDR0104 is the community affairs score; MMLG is the language score; VCPRL is a precentral cortical thickness (left) score; VCIPCL/R is an inferior parietal cortical thickness (left/right) score; VVPREL/R is a precuneus volume (left/right) score; VSSFR is a superior frontal cortical area (right) score; VSCACR is a caudal anterior cingulate area (right) score; APET.Centiloid is an amyloid PET Centiloid value; Ab42d40 is an A β 42 /A β 40 ratio; NfL is a neurofilament light chain value; and GFAP is a glial fibrillary acidic protein value.

[0220] Due to the superior performance of the model based on pTau217R, subsequent evaluations focused exclusively on this model. **FIG. 21** depicts the prediction performance seen during cross-validation extended to the test set, consistent with disclosed embodiments. Performance of the pTau217R-based model in predicting tau-PET SUVR for select cortical and Braak stage regions within the test set are shown. Performance metrics include root mean squared error (RMSE), mean absolute error (MAE), and R^2 values comparing predicted versus observed tau-PET SUVR. R^2 values for this model ranged from 0.387 to 0.577 for cortical regions and 0.378 to 0.585 for the Braak stage regions. Additionally, MAE values ranged from 0.269 to 0.347 for cortical regions and 0.238 to 0.370 for the Braak stage regions, further underlining the robustness of the model's predictions.

[0221] As depicted in **FIG. 22**, incorporating ApoE4 allelic count into the model did not improve the prediction performance, consistent with disclosed embodiments. Performance of the model utilizing pTau217R, following the integration of ApoE4 allelic count, for predicting tau-PET SUVR in select cortical regions and the six Braak stage regions within the test set are shown. Performance metrics include root mean squared error (RMSE), mean absolute error (MAE), and R^2 values, evaluating the comparison between predicted and observed tau-PET SUVR. The incorporation of ApoE4 allelic count did not yield an enhancement in prediction performance, as compared to the model without ApoE4 allelic count.

[0222] As depicted in **FIG. 23**, the performance of the model reliant on pTau217 concentration, rather than the pTau217R-based model, demonstrated a significantly inferior performance ($p < 0.05$) in predicting tau-PET SUVR in select cortical and Braak stage regions within the test set, consistent with disclosed embodiments. This discrepancy confirmed the benefit of normalizing pTau217 concentration with nonphosphorylated Tau217, yielding the ratio pTau217R. This normalization effectively mitigates inter-individual differences, highlighting its role in enhancing the model's performance.

[0223] **FIGs. 24A** and **24B** depict the differences in SUVR prediction profiles and boundaries on reliable tau-PET SUVR predictions using a model based on plasma pTau217R for the WCGM (**FIG. 24A**) and MTL (**FIG. 24B**), consistent with disclosed embodiments.

FIGs. 25A to 25D depict boundaries on reliable tau-PET SUVR predictions using a model based on plasma pTau217R for additional cortical regions beyond those shown in **FIGs. 24A** and **24B**, namely the lateral temporal region (**FIG. 25A**), frontal region (**FIG. 25B**), occipital region (**FIG. 25C**), and parietal region (**FIG. 25D**), consistent with disclosed embodiments.

FIGs. 26A to 26F depict boundaries on reliable tau-PET SUVR using a model based on plasma pTau217R for the six Braak stage regions, consistent with disclosed embodiments. In

FIGs. 24A and 24B, FIGs. 25A to 25D, and FIGs. 26A to 26F, the upper prediction limit was determined by fitting the Richards 5-parameter logistic model to the predicted versus observed tau-PET SUVR values in the test set.

[0224] **FIG. 27** depicts the range of upper limits for reliable prediction of tau-PET SUVR in the test set by the model using plasma pTau217R for select cortical regions and the six Braak stage regions. The upper limits spanned from 1.68 to 2.3 for the cortical regions and 1.42 to 2.21 for the Braak stage regions, consistent with disclosed embodiments. To contextualize these regional prediction upper limits, quartiles (1st, 2nd, and 3rd quartiles) of the observed SUVR distribution are depicted for each region, based on data pooled from 242 A β + early Alzheimer's disease subjects across training and test sets.

[0225] **FIG. 27** further depicts the SUVR threshold for tau positivity in each region, which, when coupled with the provided upper prediction limit, can serve as an indicator of the dynamic range. The distribution of tau-PET SUVR encompassed a wide range from low to intermediate to high tau levels across each region. Notably, the upper limit of SUVR prediction significantly surpassed the tau positivity threshold for cortical and Braak stage regions. The upper limit of SUVR prediction also exceeded the 3rd quartile of observed SUVR values across all regions, apart from the Braak 1-2 regions.

[0226] **FIG. 28** depicts an evaluation of tau status prediction, distinguishing between whether the predicted SUVR value falls below or above designated thresholds, consistent with disclosed embodiments. In particular, **FIG. 28** depicts the ability of the tau-PET SUVR prediction model that uses plasma pTau217R to identify subjects within the test set below or above two SUVR thresholds: one showing tau positivity and another showing a higher threshold of 1.5 SUVR. Performance metrics include the area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity, along with their respective 95% confidence intervals. Depending on the relative importance of the false negatives and false

positives in a specific context of use, the SUVR predictions can be utilized to ensure optimal sensitivity or specificity.

[0227] **FIGs. 29A to 29D** depict ROC curves and corresponding AUROC values for predicting tau positivity and a higher threshold of 1.5 SUVR in select cortical and the six Braak stage regions, consistent with disclosed embodiments. The model demonstrated accuracy in predicting the tau positivity across various brain regions, with AUROC values ranging from 0.84 to 0.92 for cortical regions and 0.85 to 0.95 for Braak stage regions. Likewise, the model demonstrated accuracy in predicting the SUVR threshold of 1.5, with AUROC values spanning from 0.85 to 0.9 for cortical regions and 0.79 to 0.87 for Braak stage regions.

[0228] **FIG. 30** depicts the results of repeating the performance evaluation described with regards to **FIG. 28** and **FIGs. 29A to 29D** for each demographic subgroup (i.e., age, sex,) and ApoE4 subgroup in the test set, consistent with disclosed embodiments. For simplicity, only WCGM and MTL regions were considered for this assessment. Prediction performance remained consistent ($p > 0.05$) across demographic and ApoE4 subgroups in MTL and WCGM, except for certain age subgroups which may be due to the small sample size ($n < 20$) in those groups ($n=14$ subjects with age ≥ 70 and MTL tau SUVR $<$ tau positivity threshold of 1.173, and $n=19$ subjects with age < 70 and WCGM tau SUVR ≥ 1.5).

[0229] A reduction in PET scans can be estimated for a scenario in which tau-PET scans are used for confirmatory and in-depth evaluations in A β + early AD patients predicted to have tau accumulation above a specific threshold of interest using blood-based biomarkers. Such an approach presents an appealing screening strategy in certain Alzheimer's disease clinical trials. It could potentially translate to real-world clinical practice, reducing patient burden while significantly lowering costs and time expenditures. However, the success of this

strategy hinges on keeping reasonably low false negative rates (FNR) and false positive rates (FPR) for tau status prediction.

[0230] In this study, the effect of the SUVR prediction models in reducing PET scans was determined by fixing sensitivities at 95%, 90% and 80%, which corresponded to tolerable false negative rates of 5%, 10%, and 20% respectively, and examining the specificity (1 minus the FPR) of the models for predicting tau status for a threshold corresponding to tau positivity and a higher threshold of 1.5 SUVR. For simplicity, this evaluation considered only the WCGM and MTL regions. The percentage reduction in PET scans reflects the model's precision in accurately identifying subjects with tau levels below the threshold, thus aligning with its specificity.

[0231] **FIG. 31** depicts a significant potential reduction in PET scans achievable by employing an SUVR prediction model using pTau217R, consistent with disclosed embodiments. In this example, the SUVR prediction model was used to exclude subjects who had a predicted tau accumulation below a specific SUVR threshold of interest while maintaining a tolerable false negative rate (FNR). Results from the MTL and WCGM regions are shown in **FIG. 31** for the sake of simplicity. In the MTL region, the percent savings reached 54.3% for the tau positivity threshold of 1.173, and 68% for an SUVR threshold of 1.5. Similarly, in the WCGM, savings of 65.4% and 43.6%, respectively, were achieved. At a 10% FNR, PET scan savings increased to 62.9% and 76% in MTL and 75% and 75.4% in WCGM for the two SUVR thresholds, respectively. Remarkably, substantial savings were achieved with a 20% FNR, reaching 88.6% and 84% in the MTL and 88.5% and 81.5% in the WCGM for the two SUVR thresholds, respectively.

[0232] In a third study, prediction models consistent with disclosed embodiments were developed for detecting whole cortical and regional tau tangles. The prediction models were trained using MK6240 tau-PET SUVR data from a clinical trial training cohort of 354

amyloid-positive early AD patients. Global and Braak stage-specific tau positivity thresholds were established using 52 cognitively unimpaired individuals from the Lantheus/Cervaux cohort through robust one-sided 95% upper confidence limits. Bayesian models were built to predict global tau positivity and tau tangles across Braak stages 1-6. Predictors included amyloid-PET CL, cognitive measures (e.g., sub-scores and composites of CDR-SB, ADAS-Cog-13, and MMSE), and plasma pTau181, alongside demographics and ApoE4 status. Prediction performance was evaluated via internal cross-validation (IV) within the training cohort and external validation (EV) using an independent ADNI cohort of 243 subjects with the Flortaucipir tracer.

[0233] Global tau positivity prediction with amyloid-PET CL achieved AUROCs of 81% (IV) and 80% (EV). This matched performance obtained by combining cognitive assessments and plasma pTau181 (AUROCs: IV 77%, EV 81%) and surpassed individual predictors. For Braak stage prediction, a separate model using amyloid-PET CL effectively distinguished earlier from later stages (Braak 0-2 vs. 3-6; AUROCs: IV 90%, EV 87%) and performed well in advanced Braak-6 stage (AUROCs: IV 79%, EV 83%). The combination of pTau181 and cognitive assessments also accurately predicted tau across Braak 0-2 versus 3-6 (AUROCs: IV 88%, EV 80%) and Braak-6 (AUROCs: IV 74%, EV 72%).

[0234] Accordingly, models consistent with disclosed embodiments can predict global tau presence and Braak-staging using amyloid-PET CL, plasma pTau181, and cognitive assessments. Validated across different tau-PET tracers, these models can enhance the efficiency of tau-PET screening, benefiting clinical and research endeavors.

[0235] In a fourth study, prediction models consistent with disclosed embodiments were developed for quantification of brain tau or A β levels. This study examined whether plasma pTau217 can predict amyloid-PET CL and predict regional tau-PET SUVR. Predictive models using plasma pTau217 predicted a continuous degree of brain amyloid over a range

suitable for early AD patients and comparable in performance to CSF A β 42/A β 40. Other predictive models using plasma pTau217 predicted tau-PET uptake.

[0236] Plasma pTau217 and amyloid beta A β 42/A β 40 were measured using immunoprecipitation-mass spectrometry. A CL prediction model was trained in a clinical trial screening cohort (TC) of 1,242 cognitively normal subjects that received 18F-NAV4694 tracer, and validated in two other clinical trial screening cohorts (VC-1 and VC-2) of 357 and 284 early AD patients respectively, with over 90% receiving florbetaben. A MK6240 tau-PET SUVR prediction model was trained and tested using a 60-40 random split of a clinical trial cohort of 242 early AD subjects. Stochastic gradient boosting was used to construct an SGB model across brain regions. Regional tau positivity thresholds were set at the robust one-sided upper 95% confidence limit in a cohort of individuals without elevated amyloid (CL<30).

[0237] **FIG. 32A** depicts the predicted amyloid-PET CL values, consistent with disclosed embodiments. The pTau217-based model best predicted up to 77 CL, with correlations of 0.81 and 0.69 in VC-1 and VC-2 respectively. The AUROC was 92% and 97% respectively in VC-1 and VC-2 for detecting CL>20, and 94% and 89% respectively for detecting CL>50. Age was selected by the models, but adding A β 42/A β 40 and ApoE4 status did not improve the predictive performance of the models.

[0238] **FIG. 32B** depicts the tau-PET uptake predicted by the pTau217-based model in the MTL, consistent with disclosed embodiments. The pTau217-based model best predicted tau-PET uptake up to SUVR of 2.1 in the medial temporal lobe, with the upper limit ranging from 1.9 to 3.2 across brain regions, and correlations ranging from 0.55 to 0.73. The AUROC was 86% for detecting whole cortical tau, 83% to 93% in the regions of Braak stages 1-6, and 93%, 83%, and 86% in Medial Temporal, Frontal, and Parietal lobes, respectively.

[0239] A fifth study investigated the effectiveness of pTau217R and $\text{A}\beta_{42}/\text{A}\beta_{40}$ in predicting continuously valued measurements of regional amyloid levels (e.g., PET CL levels) and identifying subjects with different levels of amyloid accumulation. The predictive models obtained could support clinical screen for amyloid accumulation and reduce the need for PET scans.

[0240] The fifth study developed and validated prediction models using patient data that spanned a continuum of AD, encompassing preclinical stages through early AD. The patient data was divided into three distinct cohorts. A training cohort (TC) for constructing the prediction models comprised 904 cognitively unimpaired (CUI) and early AD subjects. These subjects were aggregated from two clinical trial cohorts, ensuring a comprehensive representation of both CUI and mildly impaired individuals. The TC included 620 CUI subjects drawn from a randomized subset of the AHEAD 3-45 screening cohort (AHEAD 3-45 Study: A Study to Evaluate Efficacy and Safety of Treatment with Lecanemab in Participants with Preclinical Alzheimer's Disease and Elevated Amyloid and in Participants with Early Preclinical Alzheimer's Disease and Intermediate Amyloid; NCT04468659). The TC also included 284 amyloid-positive ($\text{A}\beta+$) early AD subjects from the ClarityAD clinical study, for whom both amyloid PET and plasma biomarker data were available (a Study to Confirm Safety and Efficacy of Lecanemab in Participants with Early Alzheimer's Disease; NCT 03887455).

[0241] A first validation cohort (VC-1) was used to assess the performance of the prediction models. The VC-1 included the remaining 622 CUI participants from the AHEAD 3-45 screening cohort.

[0242] A second validation cohort (VC-2) comprised 357 subjects evaluated for early AD for selection in clinical trials. These subjects had a suspected clinical diagnosis of MCI or mild dementia (MMSE 24 to 30, global CDR of 0.5 or 1, and impairment in delayed word recall

during screening). Approximately 46% of these subjects tested positive for A β by the visual read of PET scans, and the remainder were amyloid negative and, therefore, had cognitive impairment due to causes other than AD. These participants were drawn from the screening cohort of two identically designed clinical trials that were part of the elenbecestat phase-3 program (A Placebo-Controlled, Double-Blind, Parallel-Group, 24 Month Study with an Open-Label Extension Phase to Evaluate the Efficacy and Safety of Elenbecestat [E2609] in Subjects with Early Alzheimer's Disease; NCT02956486, MissionAD1 and NCT 03036280, MissionAD2).

[0243] Standardized uptake value ratio (SUVR) levels from amyloid PET scans, using various tracers, were converted to CL units for subjects within the training and validation cohorts. Subjects from the AHEAD 3-45 screening cohort included in the TC and VC-1 received the 18F-NAV4694 tracer. Among the ClarityAD subjects in the TC, 89% received florbetaben, while the remaining 10% and 1% received florbetapir and flutemetamol, respectively. In VC-2, 93% of subjects from MissionAD1/2 studies received florbetaben, with the remaining 7% receiving florbetapir. The interquartile range (25th to 75th percentiles) of CL distribution in TC, VC-1, and VC-2 were 2 to 74, -0.4 to 42, and -1 to 73, respectively, showing a diverse spectrum of amyloid levels across the cognitive continuum from unimpaired to early AD. The optimal CL threshold to detect the earliest amyloid accumulation has been reported to range between 15 and 30 in various studies.

[0244] Plasma biomarkers were measured using high-throughput, mass spectrometry (MS)-based assays. Blood samples were collected into K2EDTA tubes, processed into plasma, and stored at -80°C until analysis. Plasma A β 42, A β 40, pTau217, and npTau217 concentrations were quantified using liquid chromatography with tandem MS (LC-MS/MS) analytical platform. The A β 42/A β 40 and pTau217R ratios were then calculated. The pTau217R ratio was used for subsequent analyses to normalize inter-individual differences.

[0245] The distribution of age and MMSE were compared across the training and validation cohorts using the Kruskal-Wallis test, while sex and APOE status were compared using the chi-squared test.

[0246] Distinct models predicting PET CL were formulated using A β 42/A β 40 ratio alone, pTau217R alone, and a combination of both biomarkers, with demographic variables (e.g., age, sex) included in all models. BLLR and SGB were employed in the model construction. Overfitting was avoided by internally tuning the model with hold-out datasets and cross-validation. The prediction model was implemented by assembling up to 1000 decision trees with up to three-way interactions among predictors. The ranking and relative influence of each predictor in the prediction models were derived by assessing the reduction in MSE each time the predictor was used as a root node to split the decision trees in the SGB algorithm, and these were then normalized to range from 0 to 100%. Individual conditional expectation (ICE) profiles were used to investigate the relationships between predictors and outcomes.

[0247] The prediction performance of the prediction models was evaluated through 10 iterations of 10-fold cross-validation within the TC. Subsequently, the models were evaluated in the two validation cohorts, VC-1 and VC-2. The evaluation included determining the R², RMSE, and MAE for observed versus predicted PET CL. The correlation of observed versus predicted CL levels was compared between the prediction models via Hittner's test for dependent correlations. The efficacy of incorporating APOE E4 allelic count into the models was assessed within this framework.

[0248] The range of PET CL values that can be reliably predicted by the prediction models was evaluated using a Richards five-parameter generalized logistic model. The logistic model was fitted using data from CUI subjects in VC-1 and early AD subjects in VC-2. This pooled data facilitated the estimation of the CL prediction range across the spectrum from preclinical

to early AD, with the lower and upper plateau parameters of the model defining the prediction range.

[0249] The applicability of the prediction models along the disease continuum of preclinical and early AD subjects (e.g., from initial to later stages of amyloid accumulation) was assessed using VC-1 and VC-2. The prediction models were used to assess amyloid status using a wide range of CL thresholds (15, 30, 40, 50, 70, and 90). The prediction performance was summarized via receiver operating characteristic (ROC) curves and associated metrics, the area under the ROC curve (AUROC), sensitivity, and specificity. Subsequently, the AUROC values were compared between models for each CL threshold via Delong's test with Bonferroni multiplicity adjustment for further insight into their comparative performance. This prediction performance was then evaluated for each demographic (e.g., age, sex, race) and genotypic (e.g., APOE E4 status) sub-group for the 15 CL threshold to determine whether the models performed consistently across all the subgroups. The choice to focus solely on the 15 CL threshold stemmed from the significant class imbalance and limited sample sizes observed at higher thresholds within certain subgroups.

[0250] The ability of the prediction models to potentially reduce PET scan usage by identifying subjects without amyloid buildup was assessed. A sensitivity of 95% was selected, allowing for a 5% false negative rate. The specificity of the prediction models (one minus the false positive rate) was then assessed across various CL thresholds. Assuming PET scans are exclusively performed on subjects predicted to have amyloid accumulation ($A\beta+$), accurate identification of true amyloid-negative subjects by the prediction models would hypothetically result in a decrease in PET scans usage.

[0251] **FIG. 33** depicts a table summarizing patient characteristics in the study dataset, consistent with disclosed embodiments. The patient characteristics include APOE4 status, age, sex, and MMSE scores. The dataset comprised 904 subjects in the TC and 622 and 357

subjects in the two validation cohorts (VC-1 and VC-2), respectively. Significant differences between TC, VC-1 and VC-2 are indicated in the table (* indicates significance at the $p < 0.05$ level and SD indicates standard deviation). Within the training cohort, 68% ($n = 620$) were CUI, 21% ($n = 186$) exhibited mild cognitive impairment (MCI), and 11% ($n = 98$) presented with mild AD. VC-1 exclusively consisted of CUI individuals, whereas VC-2 comprised 90% ($n = 320$) MCI cases and 10% ($n = 37$) mild AD cases. The amyloid PET CL distribution spanned a wide spectrum along the disease continuum, encompassing CUI, preclinical, and early AD populations. As may be appreciated, VC-2, which includes subjects with MCI and mild AD, exhibited a significantly lower baseline MMSE compared to other cohorts ($p < 0.05$). Furthermore, distributions of APOE E4 status, age, and sex were significantly different across the three cohorts.

[0252] **FIG. 34** depicts performance of the SGB and BLLR models on the VC-1 and VC-2 datasets, consistent with disclosed embodiments. The SGB model outperformed the BLLR model. Furthermore, models that included pTau217R exhibited significantly better correlations than models that did not include pTau217R ($p < 0.001$). Prediction models that used plasma A β 42/A β 40 achieved R^2 values of 0.3 and 0.16 in VC-1 and VC-2, respectively. In contrast, prediction models that used plasma pTau217R exhibited improved predictive performance, yielding R^2 values of 0.63 and 0.66 in VC-1 and VC-2, respectively. Prediction model that used plasma A β 42/A β 40 and plasma pTau217R maintained high predictive efficacy. Similar results were apparent for other performance metrics, including MAE and RMSE. While interindividual differences were nominally mitigated through using pTau217R in place of pTau217 concentration, models relying solely on pTau217 concentration exhibited comparable predictive performance. In VC-1 and VC-2, the R^2 values were 0.61 and 0.64 respectively for the pTau217 concentration-based model and 0.64 and 0.62 respectively for the model incorporating both pTau217 concentration and A β 42/A β 40. These values were not

significantly different from those reported for pTau217R-based models in **FIG. 34** ($p > 0.05$ based on Hittner's test for comparing dependent correlations).

[0253] **FIG. 35A** depicts the percentage relative influence of predictors in an SGB model, consistent with disclosed embodiments. The SGB model included as inputs pTau217R, A β 42/A β 40, and the demographic variables age and sex. The dominant predictor was pTau217R, with more than 80% relative influence.

[0254] **FIGs. 35B** and **35C** depict ICE profiles showing individual subject-level and average outcomes, consistent with disclosed embodiments. **FIG. 35B** depicts the ICE profile of pTau217R, while **FIG. 35C** depicts the ICE profile of A β 42/A β 40. The nature of the relationship between these biomarkers versus the predicted PET CL levels is shown for each subject (in grey) and the average subject (in black) via these individual conditional expectation profiles. The prediction profile of each subject was centered by subtracting from the predicted PET CL level corresponding to the lowest value of the predictor. These profiles reveal strong non-linear relationships characterized by sigmoidal patterns. The flexibility of the SGB algorithm allows for modeling such intricate relationships without the need for explicit assumptions or predefined specifications.

[0255] **FIGs. 36A** to **36C** depict CL prediction range and observed CL values for three prediction models using different combinations of biomarker inputs, consistent with disclosed embodiments. CL prediction range was estimated using a five-parameter logistic model. The logistic model was fit using data pooled from VC-1 and VC-2. Prediction models that included pTau217R (**FIG. 36B**, demonstrating a reliable prediction range of 10.3 to 99.6 CL) and prediction models that included both pTau217R and A β 42/A β 40 ratio (**FIG. 36C**, demonstrating a reliable prediction range of 8 to 89.1 CL) exhibited greater CL prediction range than prediction models that included A β 42/A β 40 ratio but not pTau217R (**FIG. 36A**, demonstrating a reliable prediction range of 20.8 to 51.7 CL). The predicted upper plateau

was higher using the model based solely on pTau217R, but both models with pTau217R predicted CL levels across a broad spectrum along the continuum from preclinical to early AD. The CL scale is anchored to 0, the average level in young healthy people expected to have no amyloid, and 100 in people with moderate AD. Since the anchors are averages, values for individual subjects can be less than 0 or over 100.

[0256] **FIG. 37** depicts the performance of the developed prediction models for predicting amyloid status across a spectrum of CL thresholds (15 to 90 CL), consistent with disclosed embodiments. This assessment included both CUI subjects in VC-1 and early AD subjects in VC-2. Performance metrics (AUROC, sensitivity, and specificity) with 95% confidence interval are depicted for a wide range of CL thresholds. For the A β 42/A β 40-based prediction model (e.g., that included A β 42/A β 40 ratio, but not pTau217R), AUROC values ranged from 77.5% to 83.5% in VC-1 and 68.3% to 78.1% in VC-2, showing a progressive decline in performance with increasing CL thresholds. For the pTau217R-based prediction model (e.g., that included pTau217R, but not A β 42/A β 40 ratio), AUROC values ranged from 89.9% to 93.9% in VC-1 and 90.6% to 93.8% in VC-2 ($p < 0.05$), with optimal performance achieved at the 50 CL threshold. For the combined prediction model (e.g., that included both pTau217R and A β 42/A β 40), AUROC values ranged from 91.9% to 94.7% in VC-1 and 89.3% to 93.2% in VC-2. The combined prediction model exhibited improved performance over the pTau217R-based prediction model for CL thresholds of 15 to 40 in VC-1, with AUROC values increasing by up to 2.7 percentage points ($p < 0.05$). In VC-2, while modest enhancements were noted for CL thresholds of 15 and 30, these improvements did not reach statistical significance.

[0257] **FIGs. 38A to 38H** depict ROC curves and AUROC values for the different combinations of CL level and model for subjects in VC-1 and VC-2, consistent with disclosed embodiments. In **FIGs. 38A-38H**, curve 3802 corresponds to A β 42/A β 40 &

pTau217R--based prediction models, curve 3804 corresponds to pTau217R-based prediction models, and curve 3806 corresponds to A β 42/A β 40 based prediction models. The pTau217R-based prediction model predicted amyloid status for a wide range of CL thresholds. The combined prediction model exhibited significantly improved prediction for amyloid levels up to 40 CL ($p < 0.05$) in VC-1 and improved (but not significantly improved) prediction up to 30 CL in VC-2. Performance was well maintained for higher CL thresholds. In **FIG. 38A**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 92.7%, 89.9%, and 83.5%. In **FIG. 38B**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40 -based prediction models, respectively, was 94.5%, 92.1%, and 83.5%. In **FIG. 38C**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40 -based prediction models, respectively, was 94%, 93.9%, and 80.6%. In **FIG. 38D**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 94%, 93.9%, and 80.6%. **FIG. 38E**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 92.9%, 91.6%, and 78.1%. **FIG. 38F**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 93.2%, 92.7%, and 74.4%. **FIG. 38G**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 92.8%, 93.8%, and 71%. **FIG. 38H**, the AUROC for A β 42/A β 40 & pTau217R -based prediction models, pTau217R-based prediction models, and A β 42/A β 40-based prediction models, respectively, was 92.9%, 93.8%, and 68.6%.

[0258]

[0259] Including APOE $\epsilon 4$ allelic count into the A β 42/A β 40-based prediction models significantly improved ($p < 0.05$) prediction of CL levels (e.g., R^2 values increased from 0.3 to 0.35 in VC-1 and from 0.16 to 0.22 in VC-2, respectively) and prediction of amyloid status across various CL thresholds, (e.g., AUROC gains of up to 3.1 and 4.7 percentage points in VC-1 and VC-2, respectively).

[0260] Including APOE $\epsilon 4$ allelic count into the pTau217R-based prediction models significantly improved ($p < 0.05$) prediction of amyloid status in VC-1 (e.g., increased AUROC up to 2.2 percentage points) for amyloid levels up to a 70 CL threshold. While consistent trends were observed in VC-2, they lacked statistical significance across most CL thresholds.

[0261] Including APOE $\epsilon 4$ allelic count into the combined prediction models significantly improved predictive performance only in VC-2, particularly at higher amyloid levels ($CL \geq 40$), resulting in a modest increase in AUROC of less than 1 percentage point. The performance of these CL prediction models at the 15 CL threshold was evaluated across various demographic subgroups (age, sex, race, APOE $\epsilon 4$ status). The 15 CL threshold was selected due to significant class imbalances and limited sample sizes for higher thresholds within certain subgroups. Across VC-1 and VC-2, prediction performance remained consistent ($p > 0.05$) across demographic subgroups, except within VC-2, where one model demonstrates significantly different performance among racial subgroups. This discrepancy may be the result of unreliable AUROC estimates arising from the small sample size of a racial subgroup ($n = 11$ non-White subjects with $CL > 15$). Within VC-1, which had a relatively larger sample size ($n = 22$ non-White subjects with $CL > 15$), the model's performance was not significantly different across these subgroups.

[0262] **FIG. 39** depicts the predicted effects of using CL prediction models to screen subjects for confirmatory PET scans, consistent with disclosed embodiments. This scenario assumed

that the CL prediction models would be employed as a preliminary screening tool to exclude amyloid-negative subjects. Confirmatory PET scans would be conducted on subjects predicted to be amyloid-positive. This prediction was generated using the two validation cohorts (VC-1 and VC-2). The criterion for excluding subjects was selected to maintain a 5% false negative rate (95% sensitivity) across various CL thresholds. The specificity of the prediction models (1 minus the FPR) for predicting amyloid status across CL thresholds ranging from 15 to 90 was evaluated. The reduction in PET scans, expressed as a percentage, reflects the precise identification of true A β -subjects, aligning with the specificity of the model.

[0263] As shown in **FIG. 39**, screening using the A β 42/A β 40-based prediction model would potentially achieve PET scan savings ranging from 40.5% to 47.4% and 28.7% to 37.7% across different CL thresholds in VC-1 and VC-2, respectively. Screening using the pTau217R -based prediction model would potentially achieve PET scan savings of 46.9% to 77.6% and 46.8% to 81.6% in VC-1 and VC-2, respectively. Screening using the combined prediction model would potentially achieve PET scan savings of 61.7% to 78.6% and 62.7% to 70.9% in VC-1 and VC-2, respectively. At the lower thresholds of 15 and 30 CL, the combined prediction model demonstrated more substantial PET scan savings than the pTau217R -based prediction model (e.g., savings of 66% and 78.6% as compared to 46.9% and 56.6%, respectively, in VC-1, and 68.4% and 70.5% versus 46.8% and 67.1%, respectively, in VC-2).

[0264] A sixth study investigated the effectiveness of pTau217R and a β 42/a β 40 in predicting amyloid β status. The predictive models can be used for community-based screening (CBS). Prediction results can be used to select subjects for different clinical programs. The selected subjects can undergo further evaluation/confirmation via amyloid PET or CSF. Symptomatic individuals can be identified via CogState CBB assessments.

[0265] **FIG. 40** depicts patient characteristics of a cohort for building the brain A β prediction model, consistent with disclosed embodiments. The cohort included 1,236 subjects from a study-agnostic screening cohort: 784 subjects (63.4%) were amyloid negative (AN) with Amyloid-PET Centiloid (CL) < 20; 132 (10.7%) had low amyloid (AL), with CL between 20 to 40; and 320 (25.9%) had higher amyloid (AH), CL > 40. The algorithm for predicting the 3-level ordinal amyloid status (AN, AL, AH) was constructed using the Bayesian hierarchical ordinal logistic regression. The prediction model can predict 3-level ordinal status, rather than merely a binary amyloid positive or negative status.

[0266] The prediction model considered the following predictors: demographics (e.g., age, sex, and optionally BMI); plasma A β 42/A β 40 ratio; and plasma pTau217R. The effect of including either A β 42/A β 40 or pTau217R, in addition to including both biomarkers was evaluated. The plasma markers were measured using Simoa assay. Performance accuracy was first assessed via 20 iterations of 10-fold stratified cross-validation using 70% of the cohort (n=865), and further tested in the remaining 30% (n=371).

[0267] **FIG. 41** further depicts the distribution of A β 42/A β 40 ratio and pTau217R values in the training cohort, consistent with disclosed embodiments. A clearer separation of values was observed between pTau217R values than between A β 42/A β 40 values among the groups.

[0268] **FIG. 42** depicts the odds ratio and significance of each predictor in the combined biomarker prediction model, consistent with disclosed embodiments.

[0269] **FIG. 43** depicts an ROC curve and AUROC values for the combined prediction model and three different CL values, consistent with disclosed embodiments. Curve 4302 has CL <20, and AUC = .96, curve 4304 has CL20-40, and AUC = .8, and curve 4306 has CL >40 and AUC = .96.

[0270] **FIG. 44** depicts predicted amyloid β PET status broken out by demographic characteristics for a cohort of patients obtained through community-based screening,

consistent with disclosed embodiments. Such screening included performance of a blood drawn and a CogState CBB cognitive assessment. Of the 138 subjects registered, 26 out of 138 subjects were not predicted as they were screen failures and therefore a blood sample was not collected. 42 out of 112 subjects were predicted to meet CL20-40 or CL > 40 criteria and could therefore be referred to clinical trials.

[0271] The foregoing description has been presented for purposes of illustration. It is not exhaustive and is not limited to precise forms or embodiments disclosed. Modifications and adaptations of the embodiments will be apparent from consideration of the specification and practice of the disclosed embodiments. For example, the described implementations include hardware, but systems and methods consistent with the present disclosure can be implemented with hardware and software. In addition, while certain components have been described as being coupled to one another, such components may be integrated with one another or distributed in any suitable fashion.

[0272] Embodiments herein include systems, methods, and tangible non-transitory computer-readable media. The methods may be executed, at least in part for example, by at least one processor that receives instructions from a tangible non-transitory computer-readable storage medium. Similarly, systems consistent with the present disclosure may include at least one processor and memory, and the memory may be a tangible non-transitory computer-readable storage medium. As used herein, a tangible non-transitory computer-readable storage medium refers to any type of physical memory on which information or data readable by at least one processor may be stored. Examples include random access memory (RAM), read-only memory (ROM), volatile memory, non-volatile memory, hard drives, CD ROMs, DVDs, flash drives, disks, registers, caches, and any other known physical storage medium. Singular terms, such as “memory” and “computer-readable storage medium,” may additionally refer to multiple structures, such a plurality of memories or computer-readable storage media. As

referred to herein, a “memory” may comprise any type of computer-readable storage medium unless otherwise specified. A computer-readable storage medium may store instructions for execution by at least one processor, including instructions for causing the processor to perform steps or stages consistent with embodiments herein. Additionally, one or more computer-readable storage media may be utilized in implementing a computer-implemented method. The term “non-transitory computer-readable storage medium” should be understood to include tangible items and exclude carrier waves and transient signals.

[0273] Moreover, while illustrative embodiments have been described herein, the scope includes any and all embodiments having equivalent elements, modifications, omissions, combinations (e.g., of aspects across various embodiments), adaptations or alterations based on the present disclosure. The elements in the claims are to be interpreted broadly based on the language employed in the claims and not limited to examples described in the present specification or during the prosecution of the application, which examples are to be construed as nonexclusive. Further, the steps of the disclosed methods can be modified in any manner, including reordering steps or inserting or deleting steps.

[0274] The features and advantages of the disclosure are apparent from the detailed specification, and thus, it is intended that the appended claims cover all systems and methods falling within the true spirit and scope of the disclosure. As used herein, the indefinite articles “a” and “an” mean “one or more.” Similarly, the use of a plural term does not necessarily denote a plurality unless it is unambiguous in the given context. Further, since numerous modifications and variations will readily occur from studying the present disclosure, it is not desired to limit the disclosure to the exact construction and operation illustrated and described, and accordingly, all suitable modifications and equivalents may be resorted to, falling within the scope of the disclosure. Therefore, it is intended that the disclosed

embodiments and examples be considered as examples only, with a true scope of the present disclosure being indicated by the following claims and their equivalents.

[0275] The embodiments may further be described using the following clauses:

[0276] A1. A system, comprising: at least one processor and at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for predicting a degree of neurofibrillary tangles present in a human brain, the operations comprising: obtaining a machine learning model trained to predict a class label for a patient from input data of the patient, the class label being one of a set of class labels corresponding to degrees of neurofibrillary tangles present in a human brain, the input data including imaging data, cognitive function data, and/or biomarker data of the patient; generating the class label by applying the input data to the machine learning model; and providing an indication of the class label.

[0277] A2. A system, comprising: at least one processor and at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for training a machine learning model to predict a degree of neurofibrillary tangles present in the brain of a human subject, the operations comprising: obtaining classification image data for a training patient; obtaining input data for the training patient, the input data including imaging data, cognitive function data, and/or biomarker data; determining a class label for the training patient using the classification image data, the class label being one of a set of class labels corresponding to degrees of neurofibrillary tangles present in a human brain; generating a training sample that associates the class label for the training patient with the input data for the training patient; training a machine learning model using the training sample to predict the class label from the input data; and providing the trained machine learning model to enable prediction of a degree of neurofibrillary tangles present in the brain of a human subject.

- [0278] A3. The system of clause A2, wherein the patient is amyloid-positive.
- [0279] A4. The system of any one of clauses A2 to A3, wherein the classification image data comprises tau-PET imaging data.
- [0280] A5. The system of any one of clauses A1 to A4, wherein the set of class labels corresponds to Braak stages.
- [0281] A6. The system of any one of clauses A1 to A5, wherein the cognitive function data includes 14-item Alzheimer's Disease Assessment Scale (ADAS-cog-14), clinical dementia rating sum-of-boxes (CDR-SB), or mini-mental-state-examination (MMSE) data.
- [0282] A7. The system of any one of clauses A1 to A6, wherein the biomarker data comprises plasma pTau181 data.
- [0283] A8. The system of any one of clauses A1 to A7, wherein the imaging data comprises structural brain network module values or hub region values.
- [0284] A9. The system of clause A8, wherein the operations further comprise: obtaining structural MRI data; and generating the structural brain network module values or hub region values using the structural MRI data.
- [0285] A10. The system of any one of clauses A1 to A7, wherein the imaging data comprises or depends upon cortical thickness values for brain regions.
- [0286] A11. The system of any one of clauses A1 to A7, wherein the imaging data comprises or depends upon cortical thickness values for brain regions.
- [0287] A12. The system of any one of clauses A1 to A11, wherein the input data further includes patient demographics and/or ApoE4 status.
- [0288] A13. The system of any one of clauses A1 or A4 to A12, wherein the class label is used for determining whether the patient is suitable for treatment with an anti-amyloid β (A β) protofibril antibody.

[0289] A14. The system of any one of clauses A1 or A4 to A12, wherein the class label is used for monitoring treatment efficacy in the patient.

[0290] A15. The system of any one of clauses A1 or A4 to A12, wherein the class label is used to detect a decrease in the degree of neurofibrillary tangles present in the human brain.

[0291] A16. The system of any one of clauses A1 to A15, wherein the neurofibrillary tangles comprise tau tangles.

[0292] B1. A system for training a machine learning model comprising: at least one processor; and at least one non-transitory computer-readable media containing instructions that, when executed by the at least one processor, cause the system to perform operations comprising: obtaining a training dataset comprising observations for patients, each observation for a patient including: biomarker data for the patient; and tau region values, the at least one of brain tau region values based on imaging data for the patient; training, using the training dataset, a machine learning model to predict the at least one of brain tau region value based on the biomarker data; and providing the trained machine learning model.

[0293] B2. The system of clause B1, wherein: the biomarker data comprises plasma pTau217 or plasma amyloid beta (Ab)42/Ab40 biomarker data.

[0294] B3. The system of clause B1 or clause B2, wherein: the at least one of brain tau region values comprises Tau-PET cortical-to-cerebellum standardized uptake value ratio.

[0295] B4. The system of any one of clause B1 to B3, wherein: the imaging data comprises positron emission tomography data.

[0296] B5. The system of clause B4, wherein: the positron emission tomography data is acquired using a radiotracer.

[0297] B6. The system of any one of clauses B1 to B5, wherein: the machine learning model comprises a stochastic gradient boosting model.

[0298] B7. The system of any one of clauses B1 to B6, wherein: the regions include the Medial Temporal, Frontal, and Parietal lobes.

[0299] B8. A system for predicting at least one of brain amyloid or tau region values comprising: at least one processor; and at least one non-transitory computer-readable media containing instructions that, when executed by the at least one processor, cause the system to perform operations comprising: obtaining an observation for a patient, the observation including biomarker data; obtaining the trained machine learning model of any one of clauses B1 to B7; predicting at least one of brain tau region values based on the biomarker data using the trained machine learning model; and providing the predicted at least one of brain tau region values.

[0300] C1. A system, comprising at least one processor; and at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for predicting brain tau or amyloid β status, the operations comprising: obtaining a machine learning model trained to predict brain tau or amyloid β status of a patient from subject data of the patient, the subject data including biomarker data of the patient, wherein the brain tau or amyloid β status: comprises one or more continuous-valued brain tau or amyloid β levels; and/or concerns brain tau or amyloid β levels in multiple regions of the brain of the patient; generating a prediction of brain tau or amyloid β status of the patient by applying the subject data of the patient to the machine learning model; and providing the prediction of brain tau or amyloid β status of the patient.

[0301] C2. The system of clause C1, wherein the brain tau or amyloid β status concerns the brain tau levels in multiple regions of the brain of the patient.

[0302] C3. The system of clause C2, wherein the multiple regions correspond to one or more Braak stages.

- [0303] C4. The system of any one of clauses C1-C4, wherein the multiple regions comprise five or more, ten or more, twenty or more, or fifty or more regions.
- [0304] C5. The system of any one of clauses C1-C5, wherein the brain tau or amyloid β status comprises the one or more continuous-valued brain tau or amyloid β levels.
- [0305] C6. The system of clause C6, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more tau-PET SUVR levels.
- [0306] C7. The system of clause C6, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more amyloid-PET SUVR or centiloid levels.
- [0307] C8. The system of any one of clauses C1-C7, wherein the machine learning model comprises an ensemble tree-based model.
- [0308] C9. The system of clauses C8, wherein the ensemble tree-based model comprises a Stochastic Gradient Boosting model.
- [0309] C10. The system of any one of clauses C1-C9, wherein the biomarker data comprises plasma or CSF biomarker data.
- [0310] C11. The system of any one of clauses C1-C10, wherein the biomarker data comprises phosphorylated tau levels, non-phosphorylated tau levels, or functions or combinations of phosphorylated and non-phosphorylated tau levels.
- [0311] C12. The system of any one of clauses C1-C11, wherein the biomarker data comprises a pTau217 level.
- [0312] C13. The system of any one of clauses C1-C12, wherein the biomarker data comprises a pTau217/npTau217 ratio.
- [0313] C14. The system of any one of clauses C1-C13, wherein the biomarker data comprises a A β 42/A β 40 ratio.
- [0314] C15. The system of any one of clauses C1-C14, wherein the subject data further comprises demographic data.

[0315] C16. The system of any one of clauses C1-C15, wherein the subject data further comprises genomic data.

[0316] C17. The system of any one of clauses C1-C16, wherein the subject data further comprises cognitive measure data.

[0317] C18. The system of any one of clauses C1-C16, wherein the subject data does not further comprise cognitive measure data or imaging data.

[0318] C19. The system of any one of clauses C1-C17, wherein the subject data further comprises imaging data, the imaging data including structural brain network module values or hub region values.

[0319] C20. The system of clause C19, wherein the operations further comprise: obtaining structural MRI data of the patient; and generating the structural brain network module values or hub region values using the structural MRI data.

[0320] C21. The system of any one of clauses C1-C20, wherein the patient satisfies a cognitive impairment condition.

[0321] C22. The system of clause C21, wherein the cognitive impairment condition is early AD.

[0322] C23. The system of any one of clauses C1-C22, wherein the operations further comprise: providing, based on the prediction of brain tau or amyloid β status, instructions for the patient to undergo PET imaging.

[0323] C24. The system of any one of clauses C1-C23, wherein the operations further comprise: determining, at least in part based on the prediction of brain tau or amyloid β status, whether the patient is suitable for treatment with an anti-tau antibody and/or an anti-amyloid β (A β) protofibril antibody.

[0324] C25. The system of any one of clauses C1-C24, wherein the operations further comprise: using the prediction of brain tau or amyloid β status to monitor treatment efficacy in the patient.

[0325] C26. The system of any one of clauses C1-C25, wherein the prediction of brain tau or amyloid β status includes prognostic predictions of future brain tau or amyloid β status.

[0326] C27. A system, comprising: at least one processor and at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for training a machine learning model to predict brain tau or amyloid β status, the operations comprising: obtaining a brain tau or amyloid β status for a training patient; obtaining subject data for the training patient, the subject data including biomarker data; generating a training sample that associates the brain tau or amyloid β status with the subject data for the training patient; training a machine learning model using the training sample to predict the brain tau or amyloid β status from the subject data; and providing the trained machine learning model to enable prediction of the brain tau or amyloid β status.

[0327] C28. The system of clause C27, wherein the brain tau or amyloid β status concerns brain tau levels in multiple regions of the brain of the patient.

[0328] C29. The system of clause C28, wherein the multiple regions correspond to one or more Braak stages.

[0329] C30. The system of any one of clauses C28-C29, wherein the multiple regions comprise two or more regions.

[0330]

[0331] C31. The system of any one of clauses C28-C30, wherein the multiple regions comprise five or more, ten or more, twenty or more, or fifty or more regions.

- [0332] C32. The system of any one of clauses C28-C31, wherein the brain tau or amyloid β status comprises one or more continuous-valued brain tau or amyloid β levels.
- [0333] C33. The system of clauses C32, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more tau-PET SUVR levels.
- [0334] C34. The system of any one of clauses C27-C33, wherein the machine learning model comprises an ensemble tree-based model.
- [0335] C35. The system of clause 34, wherein the ensemble tree-based model comprises a Stochastic Gradient Boosting model.
- [0336] C36. The system of any one of clauses C27-C35, wherein the biomarker data comprises plasma or CSF biomarker data.
- [0337] C37. The system of any one of clauses C27-C36, wherein the biomarker data comprises phosphorylated tau levels, non-phosphorylated tau levels, or functions or combinations of phosphorylated and non-phosphorylated tau levels.
- [0338] C38. The system of any one of clauses C27-C37, wherein the biomarker data comprises a pTau217 level.
- [0339] C39. The system of any one of clauses C27-C38, wherein the biomarker data comprises a pTau217/npTau217 ratio.
- [0340] C40. The system of any one of clauses C27-C39, wherein the biomarker data comprises a A β 42/A β 40 ratio.
- [0341] C41. The system of any one of clauses C27-C40, wherein the subject data further comprises demographic data.
- [0342] C42. The system of any one of clauses C27-C41, wherein the subject data further comprises genomic data.
- [0343] C43. The system of any one of clauses C27-C42, wherein the subject data further comprises cognitive measure data.

[0344] C44. The system of any one of clauses C27-C42, wherein the subject data does not further comprise cognitive measure data or imaging data.

[0345] C45. The system of any one of clauses C27-C43, wherein the subject data further comprises imaging data, the imaging data including structural brain network module values or hub region values.

[0346] C46. The system of clause 45, wherein the operations further comprise: obtaining structural MRI data of the training patient; and generating the structural brain network module values or hub region values using the structural MRI data.

[0347] C47. The system of any one of clauses C27-C46, wherein the training patient satisfies a cognitive impairment condition.

[0348] C48. The system of clause C47, wherein the cognitive impairment condition is early AD.

[0349] C49. A method of selecting a patient for treatment with an anti-tau therapy and/or an anti-amyloid therapy, comprising identifying the patient as having or being at risk for AD according to the system of any one of clauses C1-C26, and administering the anti-tau therapy or the anti-amyloid therapy.

[0350] C50. The method of clause C49, wherein the anti-tau therapy and/or the anti-amyloid therapy comprises an anti-tau antibody and/or an anti-amyloid antibody.

[0351] C51. The method of clause C50, wherein the anti-amyloid antibody comprises lecanemab or the anti-tau antibody comprises E2814.

[0352] C52. A method of treating a patient having or suspected of having AD, comprising identifying the patient as having or being at risk for AD according to the system of any one of clauses C1-C26, and administering an anti-tau therapy and/or an anti-amyloid therapy.

[0353] C53. The method of clause C52, wherein the anti-tau therapy and/or the anti-amyloid therapy comprises an anti-tau antibody and/or an anti-amyloid antibody.

[0354] C54. The method of clause C53, wherein the anti-amyloid antibody comprises lecanemab or the anti-tau antibody comprises E2814.

[0355] C55. A method of monitoring AD treatment efficacy, comprising: measuring an elevated level and/or number of brain regions comprising neurofibrillary tangles by obtaining a level of pTau217 and applying it to a system according to any one of clauses C1-C26; and administering a therapeutic agent and repeating the measurement, wherein a reduction or delay in progression of brain regions comprising neurofibrillary tangles indicates treatment efficacy.

[0356] As used herein, unless specifically stated otherwise, the term “or” encompasses all possible combinations, except where infeasible. For example, if it is stated that a component may include A or B, then, unless specifically stated otherwise or infeasible, the component may include A, or B, or A and B. As a second example, if it is stated that a component may include A, B, or C, then, unless specifically stated otherwise or infeasible, the component may include A, or B, or C, or A and B, or A and C, or B and C, or A and B and C.

[0357] Other embodiments will be apparent from consideration of the specification and practice of the embodiments disclosed herein. It is intended that the specification and examples be considered as example only, with a true scope and spirit of the disclosed embodiments being indicated by the following claims.

What is claimed is:

1. A system, comprising:

at least one processor; and

at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for predicting brain tau or amyloid β status, the operations comprising:

obtaining a machine learning model trained to predict brain tau or amyloid β status of a patient from subject data of the patient, the subject data including biomarker data of the patient, wherein the brain tau or amyloid β status:

comprises one or more continuous-valued brain tau or amyloid β levels;

and/or concerns brain tau or amyloid β levels in multiple regions of the brain of the patient;

generating a prediction of brain tau or amyloid β status of the patient by applying the subject data of the patient to the machine learning model; and

providing the prediction of brain tau or amyloid β status of the patient.

2. The system of claim 1, wherein the brain tau or amyloid β status concerns the brain tau levels in multiple regions of the brain of the patient.

3. The system of claim 2, wherein the multiple regions correspond to one or more Braak stages.

4. The system of any one of claims 1-4, wherein the multiple regions comprise five or more, ten or more, twenty or more, or fifty or more regions.

5. The system of any one of claims 1-5, wherein the brain tau or amyloid β status comprises the one or more continuous-valued brain tau or amyloid β levels.
6. The system of claim 6, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more tau-PET SUVR levels.
7. The system of claim 6, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more amyloid-PET SUVR or centiloid levels.
8. The system of any one of claims 1-7, wherein the machine learning model comprises an ensemble tree-based model.
9. The system of claim 8, wherein the ensemble tree-based model comprises a Stochastic Gradient Boosting model.
10. The system of any one of claims 1-9, wherein the biomarker data comprises plasma or CSF biomarker data.
11. The system of any one of claims 1-10, wherein the biomarker data comprises phosphorylated tau levels, non-phosphorylated tau levels, or functions or combinations of phosphorylated and non-phosphorylated tau levels.
12. The system of any one of claims 1-11, wherein the biomarker data comprises a pTau217 level.

13. The system of any one of claims 1-12, wherein the biomarker data comprises a pTau217/npTau217 ratio.

14. The system of any one of claims 1-13, wherein the biomarker data comprises a A β 42/A β 40 ratio.

15. The system of any one of claims 1-14, wherein the subject data further comprises demographic data.

16. The system of any one of claims 1-15, wherein the subject data further comprises genomic data.

17. The system of any one of claims 1-16, wherein the subject data further comprises cognitive measure data.

18. The system of any one of claims 1-16, wherein the subject data does not further comprise cognitive measure data or imaging data.

19. The system of any one of claims 1-17, wherein the subject data further comprises imaging data, the imaging data including structural brain network module values or hub region values.

20. The system of claim 19, wherein the operations further comprise:

obtaining structural MRI data of the patient; and

generating the structural brain network module values or hub region values using the structural MRI data.

21. The system of any one of claims 1-20, wherein the patient satisfies a cognitive impairment condition.

22. The system of claim 21, wherein the cognitive impairment condition is early AD.

23. The system of any one of claims 1-22, wherein the operations further comprise:

providing, based on the prediction of brain tau or amyloid β status, instructions for the patient to undergo PET imaging.

24. The system of any one of claims 1-23, wherein the operations further comprise:

determining, at least in part based on the prediction of brain tau or amyloid β status, whether the patient is suitable for treatment with an anti-tau antibody and/or an anti-amyloid β ($A\beta$) protofibril antibody.

25. The system of any one of claims 1-24, wherein the operations further comprise:

using the prediction of brain tau or amyloid β status to monitor treatment efficacy in the patient.

26. The system of any one of claims 1-25, wherein the prediction of brain tau or amyloid β status includes prognostic predictions of future brain tau or amyloid β status.

27. A system, comprising:

at least one processor and

at least one non-transitory computer readable medium containing instructions that, when executed by the at least one processor, cause the system to perform operations for training a machine learning model to predict brain tau or amyloid β status, the operations comprising:

obtaining a brain tau or amyloid β status for a training patient;

obtaining subject data for the training patient, the subject data including biomarker data;

generating a training sample that associates the brain tau or amyloid β status with the subject data for the training patient;

training a machine learning model using the training sample to predict the brain tau or amyloid β status from the subject data; and

providing the trained machine learning model to enable prediction of the brain tau or amyloid β status.

28. The system of claim 27, wherein the brain tau or amyloid β status concerns brain tau levels in multiple regions of the brain of the patient.

29. The system of claim 28, wherein the multiple regions correspond to one or more Braak stages.

30. The system of any one of claims 28-29, wherein the multiple regions comprise two or more regions.

31. The system of any one of claims 28-30, wherein the multiple regions comprise five or more, ten or more, twenty or more, or fifty or more regions.
32. The system of any one of claims 28-31, wherein the brain tau or amyloid β status comprises one or more continuous-valued brain tau or amyloid β levels.
33. The system of claim 32, wherein the one or more continuous-valued brain tau or amyloid β levels comprise one or more tau-PET SUVR levels.
34. The system of any one of claims 27-33, wherein the machine learning model comprises an ensemble tree-based model.
35. The system of claim 34, wherein the ensemble tree-based model comprises a Stochastic Gradient Boosting model.
36. The system of any one of claims 27-35, wherein the biomarker data comprises plasma or CSF biomarker data.
37. The system of any one of claims 27-36, wherein the biomarker data comprises phosphorylated tau levels, non-phosphorylated tau levels, or functions or combinations of phosphorylated and non-phosphorylated tau levels.
38. The system of any one of claims 27-37, wherein the biomarker data comprises a pTau217 level.

39. The system of any one of claims 27-38, wherein the biomarker data comprises a pTau217/npTau217 ratio.

40. The system of any one of claims 27-39, wherein the biomarker data comprises a A β 42/A β 40 ratio.

41. The system of any one of claims 27-40, wherein the subject data further comprises demographic data.

42. The system of any one of claims 27-41, wherein the subject data further comprises genomic data.

43. The system of any one of claims 27-42, wherein the subject data further comprises cognitive measure data.

44. The system of any one of claims 27-42, wherein the subject data does not further comprise cognitive measure data or imaging data.

45. The system of any one of claims 27-43, wherein the subject data further comprises imaging data, the imaging data including structural brain network module values or hub region values.

46. The system of claim 45, wherein the operations further comprise:

obtaining structural MRI data of the training patient; and

generating the structural brain network module values or hub region values using the structural MRI data.

47. The system of any one of claims 27-46, wherein the training patient satisfies a cognitive impairment condition.

48. The system of claim 47, wherein the cognitive impairment condition is early AD.

49. A method of selecting a patient for treatment with an anti-tau therapy and/or an anti-amyloid therapy, comprising identifying the patient as having or being at risk for AD according to the system of any one of claims 1-26, and administering the anti-tau therapy or the anti-amyloid therapy.

50. The method of claim 49, wherein the anti-tau therapy and/or the anti-amyloid therapy comprises an anti-tau antibody and/or an anti-amyloid antibody.

51. The method of claim 50, wherein the anti-amyloid antibody comprises lecanemab or the anti-tau antibody comprises E2814.

52. A method of treating a patient having or suspected of having AD, comprising identifying the patient as having or being at risk for AD according to the system of any one of claims 1-26, and administering an anti-tau therapy and/or an anti-amyloid therapy.

53. The method of claim 52, wherein the anti-tau therapy and/or the anti-amyloid therapy comprises an anti-tau antibody and/or an anti-amyloid antibody.

54. The method of claim 53, wherein the anti-amyloid antibody comprises lecanemab or the anti-tau antibody comprises E2814.

55. A method of monitoring AD treatment efficacy, comprising:

measuring an elevated level and/or number of brain regions comprising neurofibrillary tangles by obtaining a level of pTau217 and applying it to a system according to any one of claims 1-26; and

administering a therapeutic agent and repeating the measurement, wherein a reduction or delay in progression of brain regions comprising neurofibrillary tangles indicates treatment efficacy.