Exponential random graph-based eXplainable Artificial Intelligence for Alzheimer disease

Nicola Amoroso,[1,2,*] Ester Pantaleo,[2,3,*] Marianna La Rocca,[2,3,†] Loredana Bellantuono,[2,4] Saverio Pascazio,[2,3] Sabina Tangaro,[2,5] Alfonso Monaco,[2,3,*] and Roberto Bellotti[2,3,*]

[1]*Dipartimento di Farmacia - Scienze del Farmaco, Università degli Studi di Bari Aldo Moro, Bari, Italy*

[2]*Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Bari, Italy*

[3]*Dipartimento Interuniversitario di Fisica "M. Merlin", Università degli Studi di Bari Aldo Moro, Bari, Italy*

[4]*Dipartimento di Biomedicina Traslazionale e Neuroscienze, Università degli Studi di Bari Aldo Moro, Bari, Italy*

[5]*Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, Bari, Italy*

(Received 18 July 2025; accepted 6 January 2026; published 20 January 2026)

The use of statistical physics models to investigate real-world networks and reveal their underlying dynamics has shown promising results and acquired increasing attention. Here, we show how exponential random-graph (ERG) models can be suitably adopted to characterize how Alzheimer's disease (AD) affects brain connectivity. Magnetic-resonance imaging (MRI) of the brain was used to define a brain connectivity network whose nodes are the different brain regions, and the links indicate the pairwise structural relationships. Based on T1-weighted MRI brain scans of 126 normal controls (NC) and 92 AD patients, ERGs were able to outline both "global" and "local" disease patterns. Our findings demonstrate that ERGs accurately highlight how AD affects brain connectivity reaching an overall classification accuracy of 0.82 ± 0.08. Besides, ERGs outline which regions of the brain are the most affected by the disease, thus proving to be a formidable instrument also to investigate the disease pathological mechanisms; more importantly, as these effects are evaluated at patient level, they can be exploited to design innovative diagnosis support systems or to provide a novel explainable framework for decision support systems. Finally, thanks to its generality, the approach proposed in this study paves the way for further applications and investigations inquiring into the use of ERGs for other diseases and different data sources or the use of alternative models.

DOI: 10.1103/dkn2-9b2t

I. INTRODUCTION

Statistical-mechanics approaches, with an established tradition in the study of many-particle physical systems, have proven to be an effective tool in the analysis of complex networks, especially for real systems, ranging from social and technological networks to biological and infrastructural ones [1–4]. In a context where networks represent sets of nodes connected by edges, statistical mechanics offers powerful tools to analyze global properties emerging from local interactions among components. Thus, it is possible to address the inherent disorder and complexity of such systems; a valuable example, that is representative of all, deals with the scaling laws and the distribution of connections among nodes which do not follow simple patterns but rather heterogeneous and nontrivial distributions, such as those described by power-law distributions [5]. By utilizing concepts such as entropy, partition functions, and phase transitions, statistical mechanics enables the modeling and understanding of key phenomena such as network robustness, percolation, epidemic spreading, and cluster formation within complex networks. The statistical approach is particularly useful in tackling the challenges posed by the vast amount of data and the nonlinear interactions typical of real networks, offering new perspectives for the analysis and prediction of the behavior of complex systems across various fields of science and technology [6–11].

Exponential random graphs (ERGs) are an established statistical mechanics tool designed to analyze and model the structure of complex networks by fitting observed network data and inferring the underlying processes driving network formation [12–15]. ERGs can account for the complex dependencies and structural patterns often observed in real-world networks, such as clustering, homophily, and the presence of specific motifs by defining a probability distribution over all possible network configurations. Alternative strategies would consist of hypothesizing one or more microscopic mechanisms underlying the actual network structures and exploit this knowledge to suitably define the networks' dynamics; however, while such approaches are known to require several refinements in order to yield accurate predictions, ERGs can lead to accurate and reliable predictions because of their rigorous probabilistic foundations [16].

By considering appropriate physical observables over an empirical network ensemble, it is possible to obtain insights into the processes shaping the networks and to determine the likelihood of observing a particular configuration, thus allowing the testing of hypotheses about network patterns and

*These authors contributed equally to this work.
†Contact author: marianna.larocca@uniba.it

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

structures of interest. This can be of particular interest when considering medical applications, especially in the field of neurodegenerative diseases, aimed at studying brain connectivity for detecting global organizational principles [17–19] and local patterns affecting specific anatomic districts or portions of them [20–23].

In recent years, some studies have explored the use of ERGs in the medical field [24–27], nevertheless, the possibility of using them to characterize Alzheimer's disease (AD) and its patterns has not been investigated, yet. The need to explore such possibility becomes urgent considering how ERGs could, in principle, enhance the understanding of the pathophysiology of AD and contribute to the development of novel diagnostic tools and therapeutic strategies aimed at preserving brain connectivity and cognitive function. Therefore, the main goals of this work are (i) the investigation of ERGs for AD modeling and the evaluation of their informative content; (ii) the possibility of using ERGs to lay the foundations of accurate and innovative explainable diagnostic frameworks for AD and, more generally, for diseases affecting brain connectivity.

II. ERGS FOR BRAIN CONNECTIVITY

For this study, a cohort of 218 subjects including 126 normal controls (NC) and 92 AD patients was investigated. Each subject underwent a $T1$-weighted magnetic- resonance imaging (MRI) brain scan which was then normalized to the $MNI152$ template and parceled in 549 rectangular boxes or patches [28,29]. Absolute Pearson's pairwise correlations ρ were measured among patches to define a weighted brain network $G(N, E, W)$ for each subject in the cohort where N is the set of nodes of cardinality $|N| = 549$, E is the set of edges, and W is the set of weights. Although various effective definitions of brain connectivity have been proposed in the literature to discriminate between AD and NC [30,31], we adopted image-intensity correlation between patches to define brain connectivity for its higher applicability in clinical practice. Indeed, it can be easily extracted from T1-weighted MRI, one of the most widely available imaging modalities in clinical practice, it is straightforward to interpret from a clinical validation perspective, and it is computationally efficient, as it does not require complex processing steps such as brain segmentation or tractography reconstruction. Less than moderate correlations ($\rho < 0.3$) were neglected, two nodes were connected with a link only if their correlation exceeded that threshold value. Since the main goal of our study was to assess differences in connectivity between groups rather than to characterize the topological organization of the networks, we considered a fixed correlation threshold more suitable than a fixed network density. While this thresholding can represent a potential issue, in the present case it should be noted that the problems associated with the use of a fixed threshold (i.e., k dependence) [32] are expected to be limited, as all subjects underwent identical acquisition protocols and image-processing pipelines. Consequently, variations in connectivity values are not driven by external or methodological factors, but rather by subject-specific phenomena (e.g., atrophic effects), which our model is designed to capture and emphasize rather than suppress. For this reason, and as correctly pointed out by van den Heuvel *et al.* [33], a density-based approach may not be optimal in the present context, as it would mitigate pathological effects that are instead responsible for the observed loss of connectivity. In other words, this approach preserves differences in connectivity, whereas a fixed network density would impose the same number of connections across subjects, potentially masking pathological alterations in global or regional connectivity. Choice of threshold value was driven by two main reasons: (i) in neuroscience, it is common to use the 0.3 value as a threshold between poor and fair correlations [34]; (ii) although the 0.3 value is widely adopted, this choice remains questionable, and for this reason, in a previous work, we investigated the impact of this choice on the informative content of brain connectivity networks [29]. We demonstrated that the informative content of brain connectivity remains stable ($\sim$5%) over a wide range of correlation values [0.2, 0.5]. Hence a binary network $G(N, E)$ was obtained, see Fig. 1 for a pictorial representation.

Two network ensembles, one for the NC and one for the AD cohorts were defined. At this point, two related research questions were addressed: whether, using ERG framework, these ensembles were distinguishable, and to what extent these differences could be diagnostically significant. These research questions were considered from two different perspectives: local and global. It is well known that AD can have both global effects on brain connectivity and local effects on different brain regions depending on several factors. Hence, to provide a model for whole-brain connectivity, a first "global" constraint was introduced on the networks' ensembles: the total number of edges m; whereas, to highlight the "local" effects of AD another constraint was specifically taken into account: the nodal degree (which evaluates how the different brain regions behave in terms of connectivity).

A. Global constraint: the total degree

In this case, a single constraint was imposed on the network ensembles and as a consequence the Hamiltonian took the simple form [14]

$$H(G) = \theta\, m(G), \tag{1}$$

with θ representing the field coupling with the total number of edges m. The expectation value $\langle m \rangle$ is

$$\langle m \rangle = \binom{|N|}{2} \frac{1}{1+e^{\theta}} \Rightarrow \theta = \ln\left(\frac{1}{p} - 1\right), \tag{2}$$

where p is the probability for each connection to appear, see Appendix A for a detailed derivation.

To determine θ, the NC and the AD cohorts were separately taken into account. Firstly, the average adjacency matrices $\langle NC \rangle$ and $\langle AD \rangle$ representing, respectively, the two cohorts were computed; then the distributions of adjacency matrix values for both cohorts were compared to identify distinct patterns or differences between the groups, see Fig. 2.

Interestingly, the average distributions of the two cohorts are indistinguishable. Therefore, it is not possible to simply discriminate patients from controls by means of the adjacency matrix values. If we count how many subjects have a specific link in their network, we find that 98% of the links are present in only 50% of the subjects overall. As these connections appeared too closely tied to the specific subject, they were

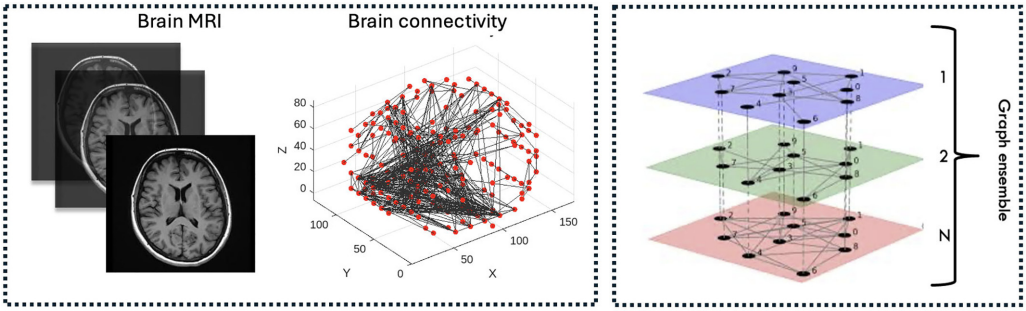


FIG. 1. Brain MRI scans collected for each patient are represented as weighted networks whose nodes are brain regions and links pairwise absolute Pearson's correlations. For each subject a graph connectivity matrix is calculated defining a graph ensemble.

not appropriate for determining general brain-connectivity patterns, so they were excluded from subsequent analyses. However, we have carried out a coupling-threshold study to see how discriminative power changes as the percentage of links being removed varies (see Appendix C).

Once these connections were removed from average matrices, empirical connection probabilities $p_{\langle NC\rangle}$ and $p_{\langle AD\rangle}$ were computed as ratios between existing links and possible links. Finally, the θ coupling was determined from Eq. (2). We ran 100 simulations by sampling 80% of subjects belonging to each cohort to estimate $\theta_{\langle NC\rangle}$ and $\theta_{\langle AD\rangle}$ along with the respective brain-connectivity Hamiltonian values; the results are presented in Fig. 3.

Coupling-field values showed a significant difference ($p < 0.01$) according to a Mann-Whitney test. This was also reflected in the Hamiltonian values, which were straightforwardly calculated according to Eq. (1). Considering that θ is a monotonically decreasing function of the connection probability p, as expected AD patients were characterized by higher coupling fields and lower brain-connectivity energies, which corresponded on average to a lower value of links with respect to NC subjects.

B. Local constraints: the degree $\vec{k}$

Given a graph $G(N, E)$ with $|N| = 549$, the set of m edges can be selected from $\binom{|N|}{2} = 150\,426$ possible edges.

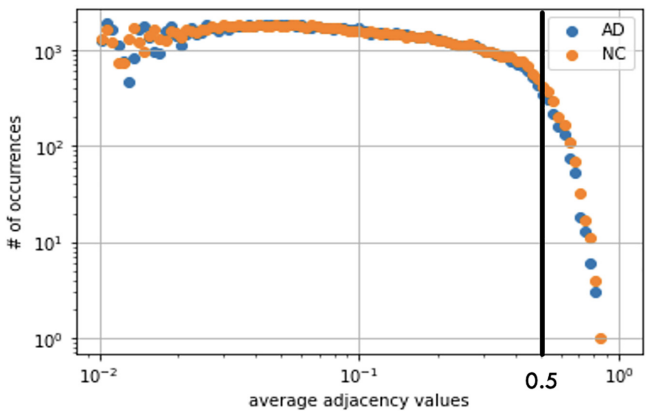


FIG. 2. The distribution of average adjacency matrices are represented for both NC and AD cohorts. Interestingly, there is substantial overlap between these distributions. Less than 5% of connections exceed 50% frequency across subjects.

Therefore, especially in the context of brain-connectivity networks, how these connections are distributed within the brain can be highly relevant. Hence, adopting a more sophisticated model (despite the increased number of parameters) can be rewarding, as it provides a more insightful representation. To this aim, following an analogous scheme to the one previously mentioned [14], the Hamiltonian H was defined as

$$H(G) = \sum_i \theta_i k_i(G), \tag{3}$$

with θ_i representing the local coupling fields (for each node i) and k_i the nodal degree. It is possible to demonstrate that

$$H(G) = \sum_{ij} \Theta_{ij} a_{ij}(G) \Rightarrow p_{ij} = \langle a_{ij}\rangle = \frac{1}{e^{\Theta_{ij}} + 1}, \tag{4}$$

where a_{ij} denotes the adjacency-matrix elements and $\Theta_{ij} = \theta_i + \theta_j$ are edge couplings. A detailed mathematical derivation is presented in Appendix B. The difference with the "global" perspective is manifest. In a global perspective, a single θ parameter accounts for brain connectivity. However, when adopting a local perspective, the number of required parameters increases dramatically, but with a key advantage: the ERG model can now be equipped with an anatomical interpretation.

As for the previous case, 100 simulations were carried out sampling 80% of subjects for both cohorts. Then, the two average matrices $\langle NC\rangle$ and $\langle AD\rangle$ were computed and, as before, connections with less than 50% rate of occurrence were discarded. Thus, the coupling fields and the resulting Hamiltonian values for both cohorts were computed. The results are shown in Fig. 4.

Even using the "local" framework, the ERG models for NC and AD showed a significant difference ($p < 0.01$) according to a Mann-Withney test. Moreover, these differences were precisely associated to network connections. This information is crucial for clinical purposes. Θ_{ij} values showed the presence of several significant differences. All nonzero differences were statistical outliers also after a Bonferroni correction; in particular, 1190 nonzero coupling differences were found out of the overall 150 426. After Bonferroni correction, 595 remained significant. The extent to which this brain-connectivity subgraph, including only the differing connections between controls and patients, is informative for AD diagnosis and useful to characterize its pathological mechanisms was also investigated.

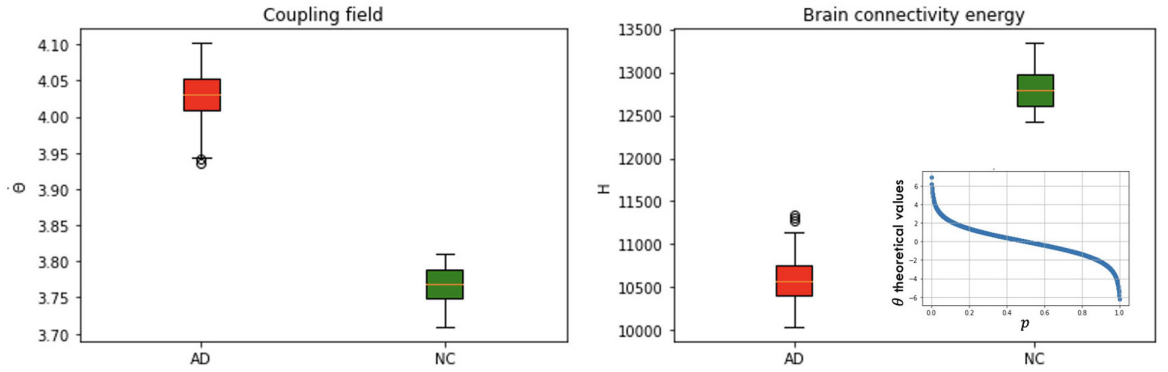


FIG. 3. Coupling-field values are significantly different in the NC and AD cohorts ($p < 0.01$). This is also reflected in the Hamiltonian values which can be interpreted as a brain-connectivity energy.

III. EVALUATING THE ERG INFORMATIVE CONTENT

While previous findings demonstrated the existence of groupwise significant differences, the effectiveness of ERG models to characterize AD for clinical purposes, i.e., at the patient level, and the impact of AD on brain connectivity deserved further studies. To address this research question and provide a quantitative evaluation of the clinical informative content provided by the ERG models, a dedicated supervised classification framework was designed.

A cross-validation procedure was implemented in which 80% of the MRI scans were randomly sampled for training and the remaining 20% for validation. The informative content provided by both the global and the local framework were assessed. At each round, the training NC and AD connectivity matrices were used to compute two average brain-connectivity matrices, namely $\langle NC \rangle$ and $\langle AD \rangle$, and the training examples were used to estimate (i) $\langle m \rangle$ for the global approach and (ii) $\langle a_{ij} \rangle$ for the local one; this process was repeated for 100 cross-validation rounds to ensure robust evaluation. Let us see in details the two cases.

For the global approach, the matrix difference $\Delta = \langle NC \rangle - \langle AD \rangle$ was computed and it was possible to outline the Δ^+ connections which were more frequent in NC subjects ($\Delta > 0$) and the Δ^- ones which were more frequent in AD patients ($\Delta < 0$). Hence, it was straightforward to define a classification score s for each subject:

$$s = \frac{\sum_{\Delta^+} \Delta_{ij}\, a_{ij}}{|\Delta^+|} - \frac{\sum_{\Delta^-} \Delta_{ij} a_{ij}}{|\Delta^-|} \tag{5}$$

where $|\Delta^+|$ and $|\Delta^-|$ denote the number of connections more frequently present in NC and AD subjects, respectively, a_{ij} are the elements of the adjacency matrix of the tested subject. The s score is composed of two terms: the first accounts for the normalized contribution of connections found only in NC subjects, and the second for those found in AD patients. Thus, if $s > 0$ the tested subject is labeled NC, otherwise AD.

The cross-validation performance was evaluated in terms of three distinct metrics: accuracy (0.65 ± 0.11), sensitivity (0.77 ± 0.13), and specificity (0.44 ± 0.17). In this context, the positive class represents NC subjects, while the negative class represents AD patients. Therefore, specificity evaluates the algorithm's ability to correctly identify AD patients, while sensitivity assesses its performance in detecting NC subjects. Although these results were far from being exceptional, they demonstrated that ERGs were able to provide a significant amount of information. More interestingly, although AD is known to affect specific brain regions, globally constrained ERGs were able to highlight disease patterns. Higher sensitivity was caused directly by the higher homogeneity characterizing NC subjects, especially compared with

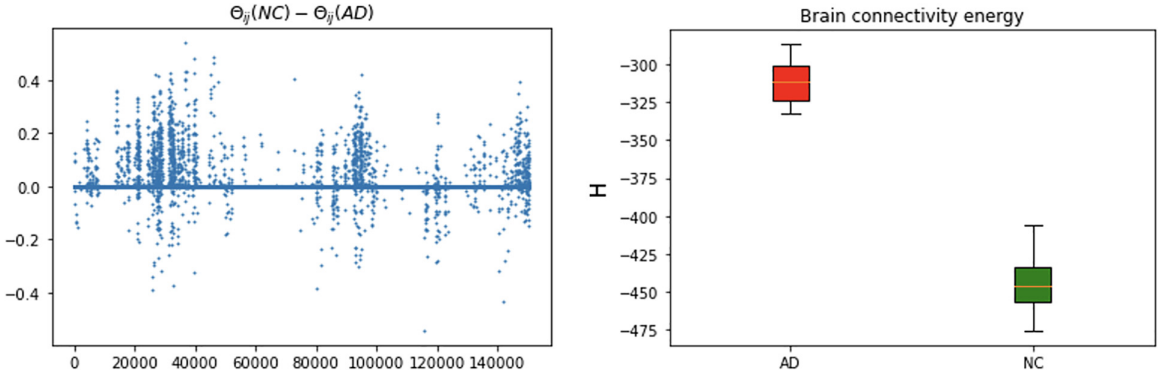


FIG. 4. Edge-couplings differences are shown in the left panel for all 150 426 possible edges, while brain connectivity energy, in terms of Hamiltonian, is shown for both cohorts in the right panel. Both couplings and Hamiltonian values show significant differences between the two cohorts.

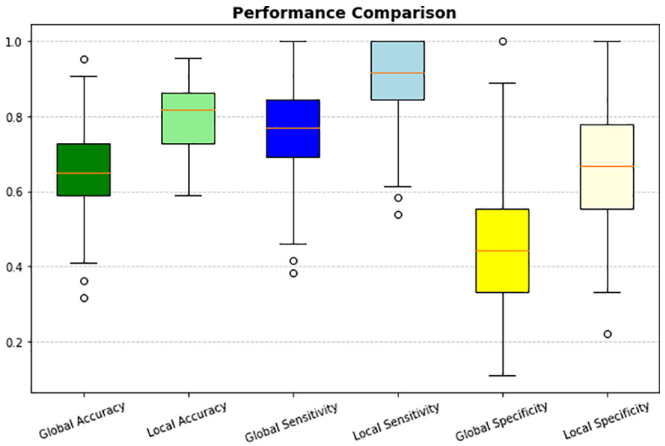


FIG. 5. The figure shows the performance of both globally and locally constrained ERG models in terms of accuracy, sensitivity, and specificity. In all three of these metrics, the local model outperforms the global one.

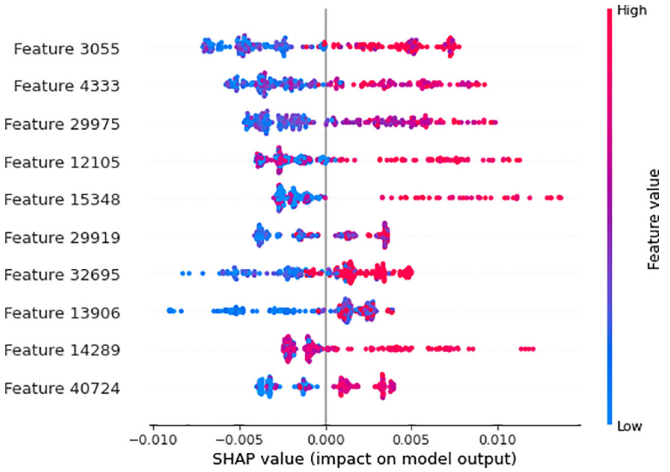


FIG. 6. The 10 most important features according to their SHAP values are shown. These features correspond to specific connections and, therefore, allow the identification of specific brain regions mostly involved in the discrimination between controls and patients.

AD patients where the disease effects on brain connectivity can yield heterogeneous patterns, see Fig. 5.

Analogously, a classification method was applied within the local framework. However, in this case, all the nonzero Δ elements fed a machine-learning model. A multivariate machine-learning approach, namely a random forest classification (RF), was adopted [35]. This choice can be motivated in two ways: (i) RF is a robust and accurate machine-learning model requiring the tuning of only two parameters, namely the number of classification trees used to grow the forest and the number of features randomly sampled at each split of the tree (for the purposes of this work the standard configuration was adopted without any tuning); (ii) RF is a state-of-the-art model widely used across various fields including medicine [36,37], remote sensing [38–40], pharmaceutical sciences [41–43], ecology [44,45] and economics [46,47], to name just a few.

Cross-validation performance was significantly enhanced, see Fig. 5. In fact, accuracy was 0.82 ± 0.08, sensitivity 0.92 ± 0.08, and specificity 0.67 ± 0.15. Again, this performance revealed that the model was more accurate in detecting NC than AD patients. More importantly, in this case it was possible to highlight which brain connections (and therefore which brain regions) were mostly responsible for the assignment of a subject to one of the two NC and AD cohorts.

IV. FROM FIELD COUPLING TO ANATOMICAL INTERPRETATION

To enhance interpretability of ERGs an eXplainable Artificial Intelligence (XAI) framework was adopted to conceive transparent explanations of how the selected brain regions influenced the classification of controls and patients. XAI is a fundamental resource for cutting-edge studies encompassing the adoption of sophisticated computational models to life sciences and, more generally, to all those domains where expert decisions must be supported by clear and transparent protocols. Hence, the main goal of XAI approaches is to transform models, often perceived as "black boxes" by non-expert practitioners, in plain and understandable classification rules. To this aim, the SHapley Additive explanation (SHAP) framework was adopted.

SHAP explanations are borrowed from game theory. The features used to feed a learning algorithm, in this case the brain connections outlined by ERGs, are interpreted as collaborative players of a game; collaborating is the key to success, which for a learning algorithm is to achieve a correct classification (of course, this should be considered in relation to the present case, but it also applies to other kind of AI tasks). Therefore, SHAP values assess how much each feature contributes to the correct classification and provides a clear and understandable way to rank all available features in terms of these contributions, what can also be considered a feature importance measure, and, more importantly, they provide subject-level explanations which "explains" how single decisions were taken by the model. By definition, the SHAP value $s(j)$ of a feature j is computed by comparing how the classification scores change by including or removing this feature from the observation $\vec{x} = (x_1, ..., x_j, ..., x_N)$ defined by all the N available features:

$$s_{\vec{x}}(j) = \sum_{S: j \in S} \frac{[p_S(\vec{x}) - p_{S/j}(\vec{x})]}{\left[|S| \times |\binom{|F|}{|S|}\right]}, \tag{6}$$

where $\times$ denotes the set cardinality, the sum is intended over all possible subsets S of features which include the feature j, and the numerator evaluates the differences between the probability values p computed including and removing the feature j from S. Results for the top 10 features are shown in Fig. 6.

Besides, the 595 connections showing significant differences between controls and patients defined a subgraph which could be easily visualized and further investigated; in particular, we evaluated which were the brain regions mostly involved within this subgraph.

To this aim, we removed all unimportant connections (in the sense that they were not related to significant differences

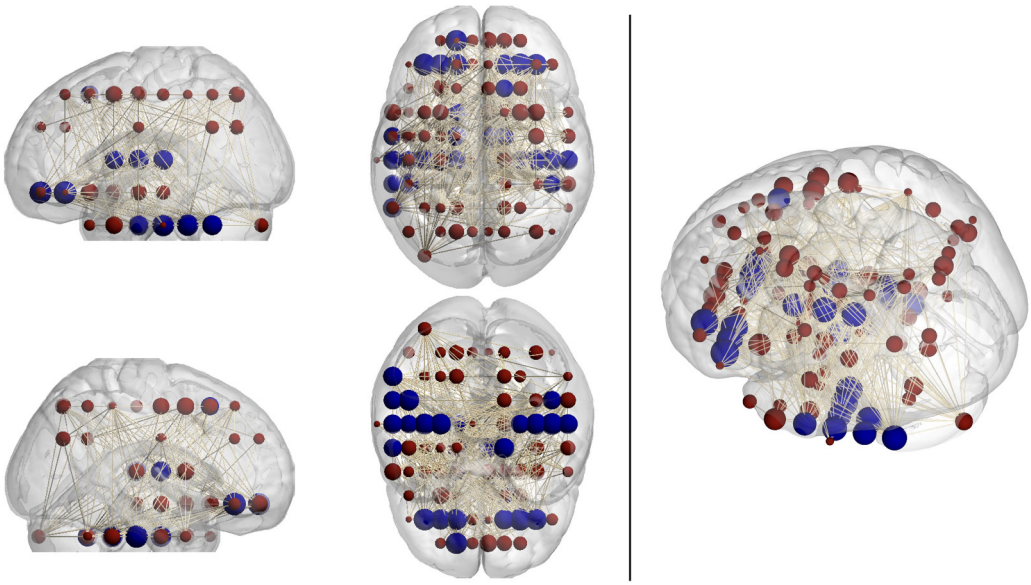


FIG. 7. Glass brain with the significant links connecting high-degree nodes (in blue) and their neighbors (in red). High-degree nodes are those with a degree greater than the 95th percentile of the degree distribution. Node size is proportional to the node degree. On the left, following chronological order, lateral view of left hemisphere, dorsal side of the brain, lateral view of right hemisphere, and ventral side of the brain are shown. On the right, the three-dimensional (3D) rendering of the glass brain is displayed.

between NC and AD cohorts) and we visualized the edges connecting high-degree nodes and their neighbors, where high-degree nodes are the nodes whose degree exceeds the 95th percentile of the degree distribution, see Fig. 7 obtained with the BrainNet Viewer [48].

Occipital lobe (posterior part of the brain) and hippocampal regions (inner anatomical districts) seemed particularly densely populated, an interesting finding given that these regions are related to cognitive functions and visual perception. Also, important nodes were located in the parietal lobe (superior part of the brain) which is associated with difficulties in performing gestures and skilled movements. These findings are confirmed by previous literature studies [49–56]. It is worth noting that the primary objective of our study was not to obtain the best possible performance, but rather to offer an innovative method based on ERG models and statistical mechanics to identify brain connectivity changes in Alzheimer's disease (AD) and to study Alzheimer's disease from a different perspective that could give new insights on processes shaping pathological brain networks. However, future research could usefully investigate the obtained subgraph, for example considering other metrics besides degree, hybrid approaches which combine global and local features, or evaluating percolation patterns and motifs. Also, the underlying geometry of such subgraph could be of interest for clinical evaluations.

For the sake of simplicity, an undirected subgraph just highlighting important connections of the initial graph (which in fact was undirected) was considered. However, taking into account the sign differences of Θ_{ij} would naturally lead to a directed graph where the sign information could be included by means of oriented links. How to interpret such model and investigate its behavior falls beyond the scope of this work, nevertheless it could represent an interesting upgrade of the proposed framework.

V. CONCLUSIONS

In this study, we have demonstrated how ERGs can effectively model brain connectivity, particularly in distinguishing between healthy individuals and those affected by Alzheimer's disease. By leveraging eXplainable Artificial Intelligence (XAI) techniques, specifically Shapley explanations, we have further identified the most influential brain connections contributing to the model's predictions. These findings provide valuable insights into the key brain regions involved in the disease, offering a data-driven approach to understanding the structural alterations associated with Alzheimer's. Our results highlight the potential of ERGs as a robust framework for brain-network analysis and emphasize the importance of interpretability in computational models applied to neurological disorders. Given the generality of the proposed framework, our approach can be extended to study other neurodegenerative diseases, such as Parkinson's disease or multiple sclerosis, where alterations in brain connectivity play a crucial role. Furthermore, ERGs could be applied to psychiatric disorders, including schizophrenia and major depressive disorder, to identify network-level biomarkers and support early diagnosis. Beyond clinical applications, this methodology could also be useful in understanding connectivity changes due to aging, cognitive decline, or brain plasticity in response to rehabilitation therapies. Future work will focus on refining these models and integrating multimodal data to enhance predictive accuracy and clinical applicability.

ACKNOWLEDGMENTS

Authors were supported by the Italian Ministry of University and Research funding within the "Budget MIUR - Dipartimenti di Eccellenza 2023 - 2027" (Law 232, 11 December 2016) - Quantum Sensing and Modelling for One-Health (QuaSiModO), CUP: H97G23000100001. Authors

were supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021 of Italian Ministry of University and Research, funded by the European Union - NextGenerationEU; Project Code CN00000013, Concession Decree No. 1031 of 17 February 2022 adopted by the Italian Ministry of University and Research, CUP: H93C22000450007, Project Title: National Centre for HPC, Big Data and Quantum Computing. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [57]. The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

N.A. designed the experiments. N.A. and M.L.R. carried out the analyses. N.A., E.P., and M.L.R. wrote the original draft. A.M. and R.B. supervised the analyses. All authors revised and approved the manuscript.

DATA AVAILABILITY

The data that support the findings of this article are openly available [57].

APPENDIX A: ERGS WITH A GLOBAL CONSTRAINT

Let us consider a graph ensemble of l graphs $\mathcal{G} = G_1, G_2, ..., G_l$ and a set of r observables $x_1, x_2, ..., x_r$. Then, it is possible to determine the expectation values of the observable $\langle x_i \rangle$. Let us consider the Gibbs entropy

$$S = -\sum_{G\in\mathcal{G}} P(G)\ln P(G) \tag{A1}$$

with the constraints

$$\sum_G P(G) = 1, \tag{A2}$$

$$\langle x_i \rangle = \sum_G x_i(G)P(G). \tag{A3}$$

Let α and θ_i denote the Lagrange multipliers, then the maximum entropy is

$$\frac{\partial}{\partial P(G)}\left[S + \alpha\left(1 - \sum_G P(G)\right) + \sum_i \theta_i\left(\langle x_i\rangle - \sum_G P(G)x_i(G)\right)\right] = 0, \tag{A4}$$

$$\rightarrow \ln P(G) + 1 + \alpha + \sum \theta_i x_i(G) = 0, \tag{A5}$$

$$\rightarrow P(G) = e^{-H(G)}/Z, \tag{A6}$$

where $Z = \sum_G e^{-H(G)}$ is the partition function and

$$H(G) = \sum_i \theta_i x_i(G) \tag{A7}$$

the Hamiltonian. Let us set as the only constraint for $\mathcal{G}$ the number of edges m. Hence, according to Eq. (A7) the resulting Hamiltonian is

$$H = \theta\, m(G). \tag{A8}$$

If a_{ij} denote the adjacency matrix elements, the partition function Z becomes

$$Z = \sum_{a_{ij}} e^{-\theta\sum_{i<j} a_{ij}} \tag{A9}$$

$$= \prod_{i<j}\sum_{a_{ij}=0}^{1} e^{-\theta a_{ij}} \tag{A10}$$

$$= \prod_{i<j}(1 + e^{-\theta}) = (1 + e^{-\theta})^{\binom{n}{2}} \tag{A11}$$

with n denoting the number of nodes. Thus, it is possible to straightforwardly calculate the expectation value

$$\langle m\rangle = \frac{1}{Z}\sum_G m e^{-H} = -\frac{1}{Z}\frac{\partial Z}{\partial\theta} = \binom{n}{2}\frac{1}{1+e^{\theta}}. \tag{A12}$$

Now, recall that the expectation value for the number of links is also given in terms of connection probability p:

$$\langle m\rangle = \binom{n}{2}p \Rightarrow \theta = \ln\left(\frac{1}{p} - 1\right). \tag{A13}$$

The θ parameter can therefore be estimated from the ensemble of observed graphs.

APPENDIX B: ERGS WITH LOCAL CONSTRAINT

Let us now consider the case in which nodal degrees k_i are used as local constraints over the ERG model. In this case, the Hamiltonian is

$$H = \sum_i \theta_i k_i = \sum_{ij}\theta_i a_{ij} = \sum_{i<j}(\theta_i + \theta_j)a_{ij} \equiv \sum_{i<j}\Theta_{ij}a_{ij}, \tag{B1}$$

where the Θ parameters take into account the contribution of each edge. Analogously to the previous global case, the partition function is

$$Z = \prod_{i<j}(1 + e^{-\Theta_{ij}}) \tag{B2}$$

and the Θ values can be estimated in terms of connection probability. However, it should be noted that in this case a different connection-probability value must be considered for each link:

$$p_{ij} = \langle a_{ij}\rangle = -\frac{1}{Z}\frac{\partial Z}{\partial\Theta_{ij}} = \frac{1}{1+e^{\Theta_{ij}}}. \tag{B3}$$

APPENDIX C: COUPLING THRESHOLD STUDY

From the study of the classification performances as a function of the coupling threshold (see Figs. 8 and 9), we can observe that the classification performances remain stable over a wide range of thresholds until the network becomes too sparse ($\sim$0.8) to have reliable metric computations. For low thresholds, when the networks are dense, performances tend to remain consistently high, however, we believe that adopting intermediate thresholds is more appropriate, as it helps avoid confounding elements from the perspective of explainability without a significant decrease in performance.

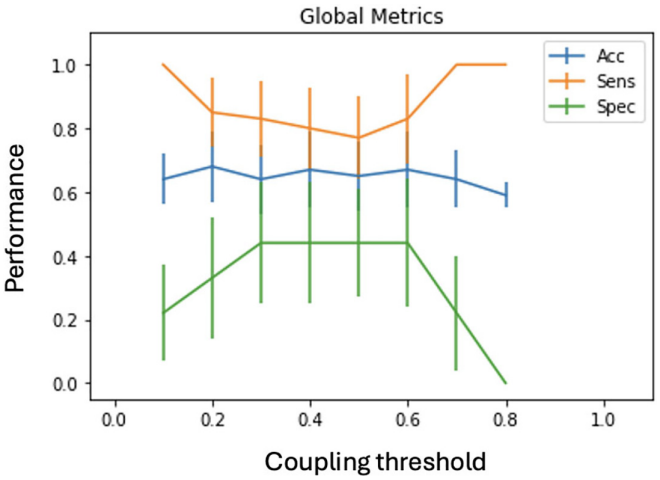


FIG. 8. Classification performances as a function of the coupling threshold of the global constraint. The lower the coupling threshold, the higher the network density

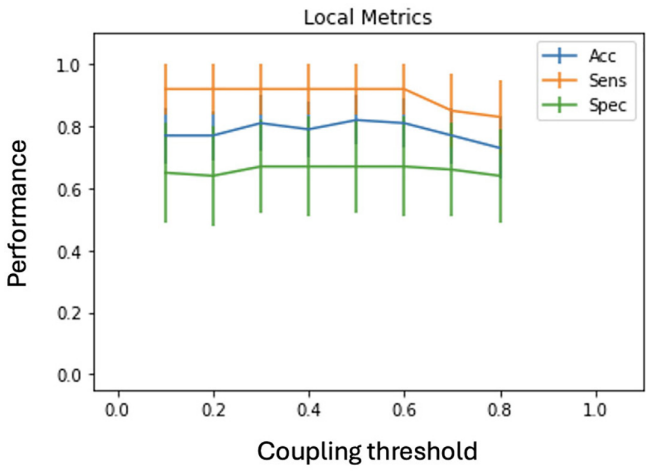


FIG. 9. Classification performances as a function of the coupling threshold of the local constraint. The lower the coupling threshold, the higher the network density.

[1] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli, The physics of financial networks, Nat. Rev. Phys. **3**, 490 (2021).

[2] J. Fan, J. Meng, J. Ludescher, X. Chen, Y. Ashkenazy, J. Kurths, S. Havlin, and H. J. Schellnhuber, Statistical physics approaches to the complex earth system, Phys. Rep. **896**, 1 (2021).

[3] M. Gosak, M. Milojević, M. Duh, K. Skok, and M. Perc, Networks behind the morphology and structural design of living systems, Phys. Life Rev. **41**, 1 (2022).

[4] G. Caldarelli, E. Arcaute, M. Barthelemy, M. Batty, C. Gershenson, D. Helbing, S. Mancuso, Y. Moreno, J. J. Ramasco, C. Rozenblat, *et al.*, The role of complexity for digital twins of cities, Nat. Comput. Sci. **3**, 374 (2023).

[5] A.-L. Barabási, Scale-free networks: A decade and beyond, Science **325**, 412 (2009).

[6] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[7] G. Bianconi, The entropy of randomized network ensembles, Europhys. Lett. **81**, 28005 (2008).

[8] E. Estrada and N. Hatano, Communicability in complex networks, Phys. Rev. E **77**, 036111 (2008).

[9] F. Radicchi, Percolation in real interdependent networks, Nat. Phys. **11**, 597 (2015).

[10] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[11] S. Boccaletti, P. De Lellis, C. Del Genio, K. Alfaro-Bittner, R. Criado, S. Jalan, and M. Romance, The structure and dynamics of networks with higher order interactions, Phys. Rep. **1018**, 1 (2023).

[12] J. Berg and M. Lässig, Correlated random networks, Phys. Rev. Lett. **89**, 228701 (2002).

[13] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Network transitivity and matrix models, Phys. Rev. E **69**, 026106 (2004).

[14] J. Park and M. E. J. Newman, Statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[15] B. A. Desmarais and S. J. Cranmer, Statistical mechanics of networks: Estimation and uncertainty, Physica A **391**, 1865 (2012).

[16] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, The statistical physics of real-world networks, Nat. Rev. Phys. **1**, 58 (2019).

[17] E. Bullmore and O. Sporns, Complex brain networks: Graph theoretical analysis of structural and functional systems, Nat. Rev. Neurosci. **10**, 186 (2009).

[18] M. Rubinov and O. Sporns, Complex network measures of brain connectivity: Uses and interpretations, NeuroImage **52**, 1059 (2010).

[19] E. Lella, N. Amoroso, A. Lombardi, T. Maggipinto, S. Tangaro, R. Bellotti, and A. D. N. Initiative, Communicability disruption in Alzheimer's disease connectivity networks, J. Complex Networks **7**, 83 (2019).

[20] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Greicius, Neurodegenerative diseases target large-scale human brain networks, Neuron **62**, 42 (2009).

[21] B. M. Tijms, A. M. Wink, W. De Haan, W. M. van der Flier, C. J. Stam, P. Scheltens, and F. Barkhof, Alzheimer's disease: Connecting findings from graph theoretical studies of brain networks, Neurobiol. Aging **34**, 2023 (2013).

[22] N. Amoroso, M. La Rocca, A. Monaco, R. Bellotti, and S. Tangaro, Complex networks reveal early MRI markers of Parkinson's disease, Med. Image Anal. **48**, 12 (2018).

[23] J. W. Vogel, N. Corriveau-Lecavalier, N. Franzmeier, J. B. Pereira, J. A. Brown, A. Maass, H. Botha, W. W. Seeley, D. S. Bassett, D. T. Jones, *et al.*, Connectome-based modeling of neurodegenerative diseases: Towards precision medicine and mechanistic insight, Nat. Rev. Neurosci. **24**, 620 (2023).

[24] S. L. Simpson, M. N. Moussa, and P. J. Laurienti, An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks, NeuroImage **60**, 1117 (2012).

[25] M. R. Sinke, R. M. Dijkhuizen, A. Caimo, C. J. Stam, and W. M. Otte, Bayesian exponential random graph modeling of

whole-brain structural networks across lifespan, NeuroImage **135**, 79 (2016).
[26] B. C. L. Lehmann, R. N. Henson, L. Geerligs, and S. R. White, Characterising group-level brain connectivity: A framework using Bayesian exponential random graph models, NeuroImage **225**, 117480 (2021).
[27] Y. Fan and S. R. White, Review of weighted exponential random graph models frameworks applied to neuroimaging, Stat. Med. **43**, 3881 (2024).
[28] M. La Rocca, N. Amoroso, A. Monaco, R. Bellotti, S. Tangaro, and Alzheimer's Disease Neuroimaging Initiative, A novel approach to brain connectivity reveals early structural changes in Alzheimer's disease, Physiol. Meas. **39**, 074005 (2018).
[29] N. Amoroso, M. La Rocca, S. Bruno, T. Maggipinto, A. Monaco, R. Bellotti, and S. Tangaro, Multiplex networks for early diagnosis of Alzheimer's disease, Front. Aging Neurosci. **10**, 365 (2018).
[30] A. Marcisz, A. D. N. Initiative, and J. Polanska, Can T1-weighted magnetic resonance imaging significantly improve mini-mental state examination-based distinguishing between mild cognitive impairment and early-stage Alzheimer's disease? J. Alzheimer's Dis. **92**, 941 (2023).
[31] D. Putcha, Y. Katsumi, M. Brickhouse, R. Flaherty, D. H. Salat, A. Touroutoglou, and B. C. Dickerson, Gray to white matter signal ratio as a novel biomarker of neurodegeneration in Alzheimer's disease, NeuroImage: Clinical **37**, 103303 (2023).
[32] B. C. Van Wijk, C. J. Stam, and A. Daffertshofer, Comparing brain networks of different size and connectivity density using graph theory, PLoS One **5**, e13701 (2010).
[33] M. P. van den Heuvel, S. C. de Lange, A. Zalesky, C. Seguin, B. T. T. Yeo, and R. Schmidt, Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations, NeuroImage **152**, 437 (2017).
[34] H. Akoglu, User's guide to correlation coefficients, Turk. J. Emerg. Med. **18**, 91 (2018).
[35] L. Breiman, Random forests, Mach. Learn. **45**, 5 (2001).
[36] R. Díaz-Uriarte and S. Alvarez de Andrés, Gene selection and classification of microarray data using random forest, BMC Bioinf. **7**, 1 (2006).
[37] X. Chen and H. Ishwaran, Random forests for genomic data analysis, Genomics **99**, 323 (2012).
[38] M. Pal, Random forest classifier for remote sensing classification, Int. J. Remote Sens. **26**, 217 (2005).
[39] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, Random forests for land cover classification, Pattern Recognit. Lett. **27**, 294 (2006).
[40] R. Cilli, A. Monaco, N. Amoroso, A. Tateo, S. Tangaro, and R. Bellotti, Machine learning for cloud detection of globally distributed Sentinel-2 images, Remote Sens. **12**, 2355 (2020).
[41] G. Riddick, H. Song, S. Ahn, J. Walling, D. Borges-Rivera, W. Zhang, and H. A. Fine, Predicting *in vitro* drug sensitivity using random forests, Bioinformatics **27**, 220 (2011).
[42] G. Cano, J. Garcia-Rodriguez, A. Garcia-Garcia, H. Perez-Sanchez, J. A. Benediktsson, A. Thapa, and A. Barr, Automatic selection of molecular descriptors using random forest: Application to drug discovery, Expert Syst. Appl. **72**, 151 (2017).
[43] M. V. Togo, F. Mastrolorito, F. Ciriaco, D. Trisciuzzi, A. R. Tondo, N. Gambacorta, L. Bellantuono, A. Monaco, F. Leonetti, R. Bellotti, *et al.*, TIRESIA: An explainable artificial intelligence platform for predicting developmental toxicity, J. Chem. Inf. Model. **63**, 56 (2023).
[44] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, Random forests for classification in ecology, Ecology **88**, 2783 (2007).
[45] R. Feng, H.-j. Zheng, H. Gao, A.-r. Zhang, C. Huang, J.-x. Zhang, K. Luo, and J.-r. Fan, Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China, J. Cleaner Prod. **231**, 1005 (2019).
[46] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, An ensemble random forest algorithm for insurance big data analysis, IEEE Access **5**, 16568 (2017).
[47] S. Islam and S. H. Amin, Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques, Journal of Big Data **7**, 65 (2020).
[48] M. Xia, J. Wang, and Y. He, Brainnet viewer: A network visualization tool for human brain connectomics, PLoS One **8**, e68910 (2013).
[49] J. Aharon-Peretz, O. Israel, D. Goldsher, and A. Peretz, Posterior cortical atrophy variants of Alzheimer's disease, Dementia Geriatr. Cognit. Disord. **10**, 483 (1999).
[50] L. G. Apostolova, R. A. Dutton, I. D. Dinov, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson, Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps, Arch. Neurol. **63**, 693 (2006).
[51] M. R. Sabuncu, R. S. Desikan, J. Sepulcre, B. T. T. Yeo, H. Liu, N. J. Schmansky, M. Reuter, M. W. Weiner, R. L. Buckner, R. A. Sperling, *et al.*, The dynamics of cortical and hippocampal atrophy in Alzheimer disease, Arch. Neurol. **68**, 1040 (2011).
[52] H. I. Jacobs, M. P. Van Boxtel, H. B. Uylings, E. H. Gronenschild, F. R. Verhey, and J. Jolles, Atrophy of the parietal lobe in preclinical dementia, Brain Cognit. **75**, 154 (2011).
[53] S. Neufang, A. Akhrif, V. Riedl, H. Förstl, A. Kurz, C. Zimmer, C. Sorg, and A. M. Wohlschläger, Disconnection of frontal and parietal areas contributes to impaired attention in very early Alzheimer's disease, J. Alzheimers Dis. **25**, 309 (2011).
[54] H. I. L. Jacobs, M. P. J. Van Boxtel, J. Jolles, F. R. J. Verhey, and H. B. M. Uylings, Parietal cortex matters in Alzheimer's disease: An overview of structural, functional and metabolic findings, Neurosci. Biobehav. Rev. **36**, 297 (2012).
[55] R. Migliaccio, F. Agosta, K. L. Possin, E. Canu, M. Filippi, G. D. Rabinovici, H. J. Rosen, B. L. Miller, and M. L. Gorno-Tempini, Mapping the progression of atrophy in early- and late-onset Alzheimer's disease, J. Alzheimer's Dis. **46**, 351 (2015).
[56] N. Amoroso, M. L. Rocca, R. Bellotti, A. Fanizzi, A. Monaco, and S. Tangaro, Alzheimer's disease diagnosis based on the hippocampal unified multi-atlas network (HUMAN) algorithm, Biomed. Eng. Online **17**, 1 (2018).
[57] Alzheimer's Disease Neuroimaging Initiative, Alzheimer's disease neuroimaging initiative (ADNI), https://adni.loni.usc.edu (2004–), data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, accessed: 6 Jan 2026, https://go.aps.org/das.