Explainable Reinforcement Learning for Alzheimer's Disease Progression Prediction

Raja Farrukh Ali
Kansas State University
rfali@ksu.edu

Ayesha Farooq
Kansas State University
ayeshafarooq@ksu.edu

Emmanuel Adeniji
Kansas State University
adeniji@ksu.edu

John Woods
Kansas State University
jwoods03@ksu.edu

Vinny Sun
Kansas State University
vinnysun1@ksu.edu

William Hsu
Kansas State University
bhsu@ksu.edu

Abstract

We present a novel application of SHAP (SHapley Additive exPlanations) to enhance the interpretability of Reinforcement Learning (RL) models used for Alzheimer's Disease (AD) progression prediction. Leveraging RL's predictive capabilities on a subset of the ADNI dataset, we employ SHAP to explain the model's decision-making process. Our approach provides detailed insights into the key factors influencing AD progression predictions, offering both global and individual, patient-level interpretability. By bridging the gap between predictive power and transparency, our work is a step towards empowering clinicians and researchers to gain a deeper understanding of AD progression and facilitate more informed decision-making in AD-related research and patient care. To encourage further exploration, we open-source our codebase [1].

1 Introduction

Alzheimer's disease (AD) is a progressive, irreversible, neurodegenerative disease that affects millions of individuals worldwide [1]. AD is characterized by the progressive shrinking of the brain and the eventual death of neurons. It causes memory and language deterioration, cognitive deficits, and impairments in judgment as well as communication. Understanding the factors driving AD progression and developing accurate prediction models are crucial for early diagnosis, intervention, and improved patient outcomes [2].

While machine learning models have shown promise in predicting AD progression, their lack of interpretability poses a significant challenge [3]. Explainability not only helps in understanding the factors that contribute to AD progression but also enables clinicians to make informed decisions regarding patient care [4]. Interpretability in supervised machine learning, where models are trained on labeled data, differs from unsupervised machine learning, where models learn patterns from unlabeled data. In supervised learning, interpretability focuses on understanding the relationship between input features and the model's predictions. However, in reinforcement learning (RL), interpretability becomes more challenging due to the inherent sequential decision-making nature of the problem, as well as the traditionally large state and action spaces associated with RL environments [5].

In this work, we apply SHAP (SHapley Additive exPlanations) [6], a post-hoc explanation method, to enhance the interpretability of an RL model for AD progression prediction. SHAP provides a

[1] https://github.com/rfali/xrlad

XAI in Action: Past, Present, and Future Applications @ NeurIPS 2023.

method of assigning feature importance scores, quantifying the contribution of each feature to the RL model's output (i.e. actions). By leveraging SHAP, we gain valuable insights into the relative importance of different features, shedding light on the factors driving AD progression in a more interpretable manner.

2 Related Work

2.1 Modeling Alzheimer's Disease Progression

Research into modeling the progression of Alzheimer's disease can be generally classified into mechanistic models and data-driven models. Mechanistic models rely on domain knowledge to encode relationships among variables through algebraic and/or differential equations [7–10]. Data-driven models encompass a spectrum of techniques, including Bayesian models [11], event-based models [12], mixed-effects models [13, 14], and machine learning models [15–17]. These models leverage bio-marker data to establish connections between disease pathology, region size, cognitive function, and demographic factors. See [18, 19] for recent literature reviews on the use of AI/ML for AD diagnosis. These techniques exhibit a relatively low reliance on domain-specific knowledge and are effective for short-term forecasting. More recently, [20] combined mechanistic and reinforcement learning to propose a hybrid model that leverages RL's ability to model this task as a sequential decision making problem, and predict AD's progression over years based on baseline imaging/cognition data, as well as demographic features.

2.2 Explainable RL (XRL)

In response to the call for transparency in AI models, the domain of Explainable AI (XAI) has witnessed significant momentum in recent years. Presently, the majority of XAI research has been concentrated on tasks such as classification, however there has been a surge in research dedicated to Explainable Reinforcement Learning (XRL) [21]. XRL research can be categorized in various ways. The most prominent method of categorization splits XRL methods into (a) transparent methods and (b) post-hoc explainability methods [22]. Transparent methods include RL models that can be explained by themselves, whereas Post-Hoc explainability methods provide explanations of RL algorithms after the models have been trained. Our work uses SHAP, which is as a post-hoc explainability method with interaction data to explain the RL model's predictions [23]. SHAP uses the magnitude of influence from each variable in the environment after training to quantify the interactions between and contributions of each variable towards the final prediction of the model.

Another taxonomy of XRL methods categorizes methods as being one of (a) Feature Importance - FI (b) Learning Process and MDP - LPM and (c) Policy Level - PL [24]. FI explanations describe the reasoning behind taking an action, LPM explanations describe the most learning-influential experiences of the model, and PL explanations summarize the long-term behavior of the model. SHAP falls within the "Directly Generate Explanation" subcategory within FI. SHAP generates an explanation after training from a non-interpretable policy enabling understanding of the factors that influence a model towards its final predictions. Similarly, from [25], SHAP falls into the model-explaining and explanation-generating category. This category describes methods that generate explanations from the model without being explicitly self-explainable. Finally, while SHAP is a very popular library used in XAI research, it has seen very limited use in RL applications [26, 27] because of the traditionally large state and action spaces associated with RL environments. Thus, we hope our work will increase the usage of such post-hoc explanation tool in the growing field of XRL.

3 RL Model

In this section, we provide a brief overview of the RL model used to predict AD progression. The model, presented in [20], leverages domain knowledge and RL to establish causal relationships between various factors involved in AD progression. We first describe some of the important factors used for AD diagnosis. Amyloid beta (Aβ), measured using florbetapir-PET scans, propagates between brain regions, influencing brain structure (measured via MRI), activity (measured via

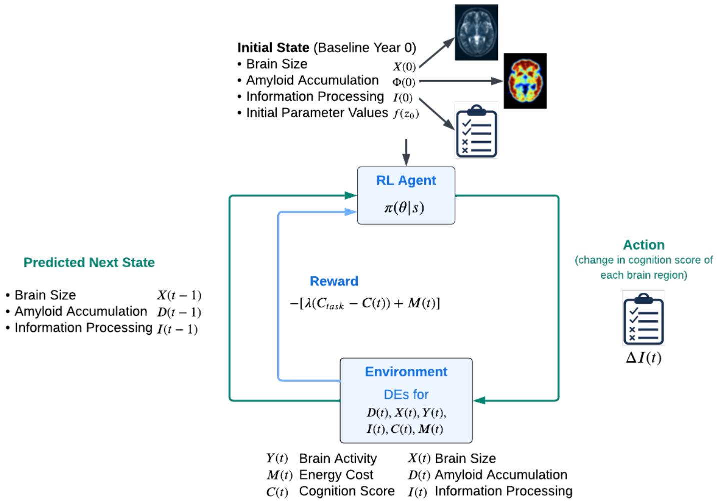

Initial State (Baseline Year 0)
• Brain Size $X(0)$
• Amyloid Accumulation $\Phi(0)$
• Information Processing $I(0)$
• Initial Parameter Values $f(z_0)$
RL Agent
$\pi(\theta|s)$
Predicted Next State
• Brain Size $X(t-1)$
• Amyloid Accumulation $D(t-1)$
• Information Processing $I(t-1)$
Reward
$-[\lambda(C_{task} - C(t)) + M(t)]$
Action
(change in cognition score of each brain region)
$\Delta I(t)$
Environment
DEs for
$D(t), X(t), Y(t),$
$I(t), C(t), M(t)$
$Y(t)$ Brain Activity
$M(t)$ Energy Cost
$C(t)$ Cognition Score
$X(t)$ Brain Size
$D(t)$ Amyloid Accumulation
$I(t)$ Information Processing

Figure 1: The RL Model

fMRI), and cognition (measured through tests like Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale (ADAS11, ADAS13)). The model defines a hypothetical variable, C_{task}, which represents cognitive demand, and impacts brain activity. Brain activity, in turn, affects cognition and contributes to neurodegeneration. The model also considers the energetic cost associated with brain activity, which can further contribute to neurodegeneration. The model defines these relationships using appropriate sets of differential equations (DEs), the variables of which are defined below, whereas the DEs are provided at Appendix A. We refer the interested reader to [20] for a detailed overview of the model.

The brain is represented as a graph $G_S = (V, E)$, where a node $v \in V$ represents a brain region, and an edge $e \in E$ represents a tract. Let $X_v(t)$ denote the size of a brain region $v \in V$ at time t, and $X(t) = [X_1(t), X_2(t), ..., X_{|V|}(t)]$. $D_v(t)$ is the instantaneous amyloid accumulation in region $v \in V$ at time t. The total amount of amyloid in a region is $\varphi_v(t)$. $Y_v(t)$ denotes the activity in region $v \in V$ in support of cognition $C(t)$ at time t. Although cognition, brain size (X_v), and activity (Y_v) are related, the exact relationship among them is unknown and cannot be easily learned from limited data. The energetic cost $M(t)$ represents the brain's energy consumption, which is proportional to its overall activity $Y_v(t)$ and serves as a cost associated with supporting cognition. The model employed investigates two regions of interest, the hippocampus (HC) and prefrontal cortex (PFC).

DEs provide relationships between some, but not all, factors relevant to AD. To address missing relationships, the model formulates an optimization problem, which it solves using reinforcement learning (RL). Fig 1 explains the working of the RL model. The environment is represented as a simulator that encompasses the equations governing various factors, including $D(t)$, $\phi(t)$, $X(t)$, $Y(t)$, $I(t)$, $C(t)$, and $M(t)$. The state at time t, denoted as $S(t)$, comprises the current sizes of the brain regions, $X(t)$, the amyloid accumulation $D(t)$ and the information processed by each region at the previous time step, $I(t-1)$. The action at time t, denoted as $A(t)$, is an element of the action space A and specifies the change in information processed by each brain region from the previous time point, i.e., $\Delta I_v(t) \in \mathbb{R}$ for all $v \in V$.

The RL agent aims to balance the trade-off between two competing criteria: (i) minimizing the discrepancy between the cognitive demand C_{task} and the actual cognition $C(t)$ and (ii) minimizing the cost $M(t)$ associated with supporting cognition. The reward $R(t)$ at time t is defined as follows, where λ is a parameter controlling the trade-off between the mismatch and the cost, and the agent's goal is to maximize this reward given by:

$$R(t) = -\left[\lambda(C_{\text{task}} - C(t)) + M(t)\right] \tag{1}$$

This model predicts disease progression by considering the interactions between these DEs, effectively creating a simulator, and the actions taken by the RL agent.

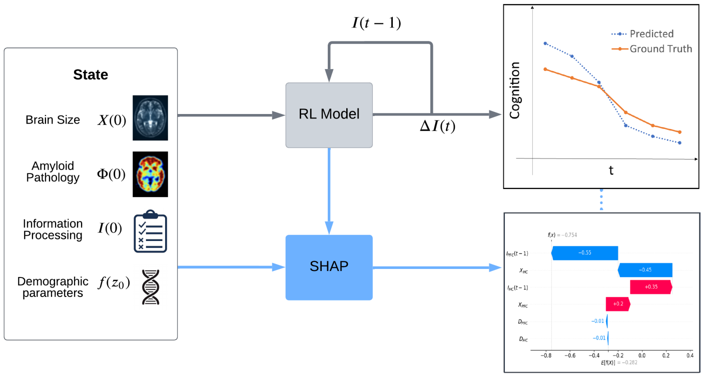


Figure 2: Using SHAP to explain the RL Model's predictions

4 XRL: Using SHAP to explain RL model's predictions

4.1 Shapley Values

Shapley values [28], rooted in cooperative game theory, assign a value to each player based on their marginal contribution to different coalitions (subsets of players) to fairly allocate the total payoff of the game to each player. The Shapley value of a player is based on their marginal contributions to all possible combinations in which they participate. Mathematically, the Shapley value for a player i in a game with N players is defined as

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \tag{2}$$

where $\phi_i(v)$ is the Shapley value of player i, $v(S)$ is the value of the coalition S, $|N|$ is the total number of players, and $|S|$ is the number of players in coalition S.

4.2 SHAP

Lundberg and Lee [6] build on Shapley values and its extensions to develop a model-agnostic interpretability framework called SHAP (SHapley Additive exPlanations), offering a robust and coherent approach to interpreting model predictions. From Eq 2, $\phi_i(v)$ is the Shapley value of a specific feature in the model, $v(S)$ is the prediction made by the RL model for a specific set of features, $|N|$ is the total number of the input features, and S ranges over all possible coalitions excluding feature i, which signifies that when calculating the Shapley value for a specific feature i, all possible combinations of features (excluding feature i) are considered. The value of S changes as different subsets of features are examined. This way SHAP values attribute a model's prediction to distinct features, explaining their influence on the model's output. Specifically, the SHAP framework can be used to achieve two types of interpretability. Global interpretability: by aggregating SHAP values computed for each individual instance across the entire dataset, the framework provides a comprehensive perspective on the behavior of the model in predicting AD across a diverse spectrum of cases. This can help to identify consistent features that significantly influence predictions. Local interpretability: by delving into the process of individual predictions for AD, considering the unique impact of each input feature, the SHAP framework can provide microscopic insights into the rationale behind specific predictions. This is pivotal for understanding the model's decision-making on a case-by-case basis, shedding light on the prominent factors guiding predictions for individual patients. Fig 2 presents our complete experimental setup, where we input the RL model as well as the State space $S(t)$ to the SHAP library and generate global (for all patient data in test split) and local explanations (predictions for each patient).

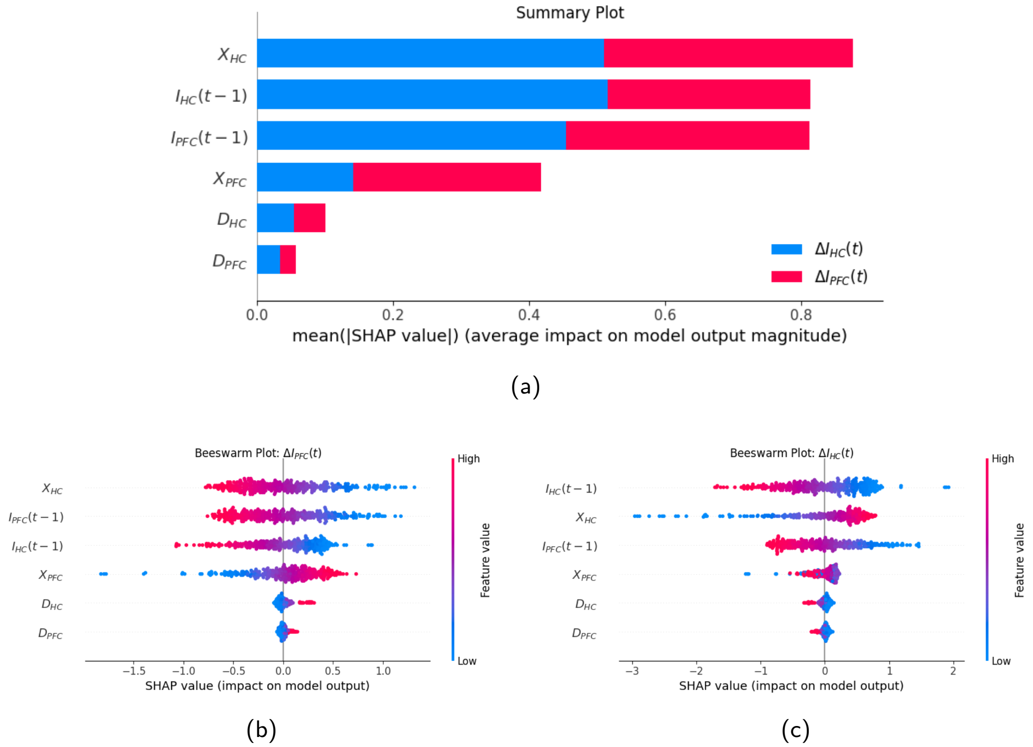


Figure 3: SHAP Plots for all patient predictions (a) Summary Bar Plot for Information processing of PFC region $I_{PFC}(t)$ and HC region $I_{HC}(t)$, (b) Beeswarm Plot for Information processing of PFC region $I_{PFC}(t)$, (c) Beeswarm Plot for Information processing of HC region $I_{HC}(t)$

4.3 Global Explanations

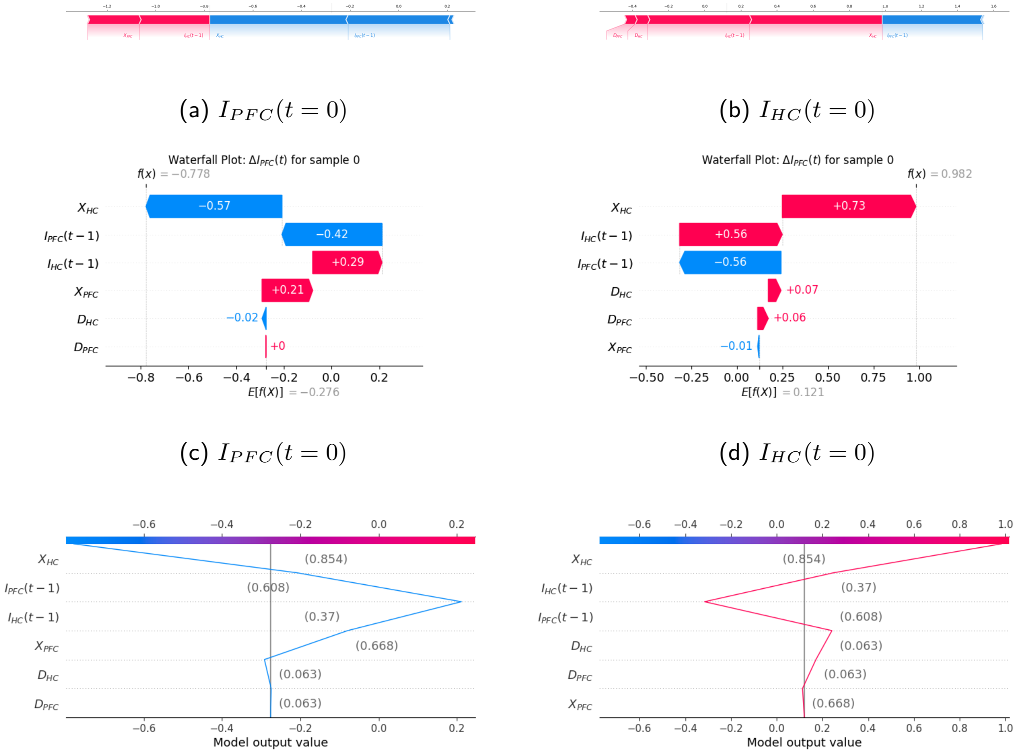
Global explanations are visualized using bar and beeswarm SHAP plots (Fig 3). The bar plot ranks features by mean absolute SHAP value for information processing in the Prefrontal Cortex region $(I_{PFC}(t=0))$ and the Hippocampus region $(I_{HC}(t=0))$. Features with higher mean absolute values are placed at the top, indicating their greater influence. The beeswarm plot assigns distinctive colors to sample values (red high, blue low), illustrating how high and low feature values impact the model's outcome.

4.4 Local Explanations

The SHAP plots used for local explanations were decision plots, waterfall plots, and force plots (Fig 4). These plots display how individual features affect the model behavior for a single sample. For the decision and waterfall plots, the expected value for all samples is displayed at the bottom. Each feature is then added to the plot starting with the least important feature. Each feature has a SHAP value that pushes the expected value positively or negatively. Once all features are added, the plot reaches the actual predicted value for the particular sample. This value is displayed at the top of the plot. Features pushing the prediction higher are shown in red, whereas those pushing the prediction lower are in blue. The force plot serves the same function as the decision and waterfall plot but the plot is visualized along the x-axis.

5 Results

In this paper, a set of SHAP plots were generated including a bar plot, a beeswarm plot, a waterfall plot, a force plot, and a decision plot. The primary aim of this section is to examine these plots, extracting information about the model's behavior, and analyzing if the model's behavior aligns with the current research within the field of Alzheimer's disease (AD).



(a) $I_{PFC}(t=0)$

(b) $I_{HC}(t=0)$

(c) $I_{PFC}(t=0)$

(d) $I_{HC}(t=0)$

(e) $I_{PFC}(t=0)$

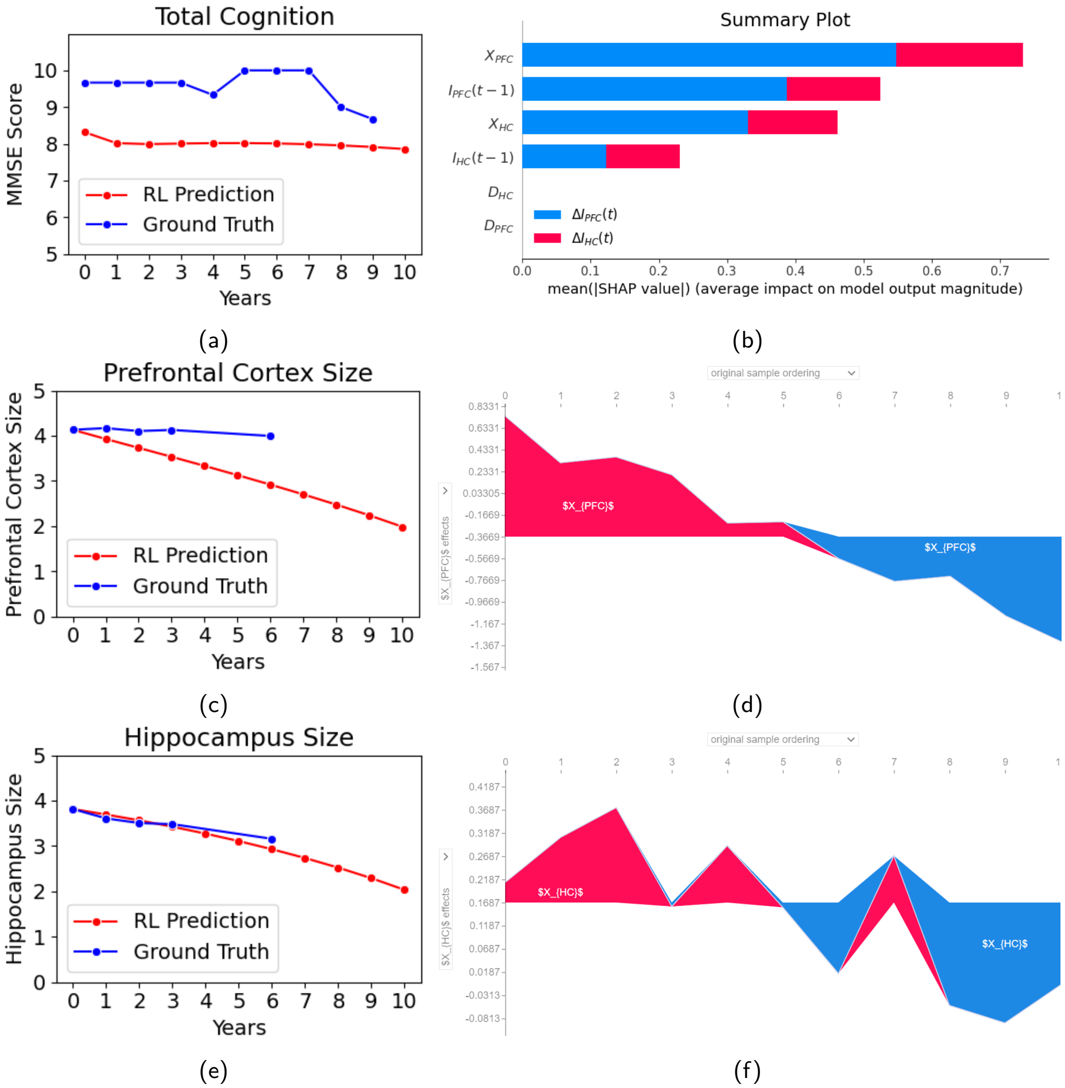
(f) $I_{HC}(t=0)$

Figure 4: SHAP Plots for a single prediction (a) and (b) Waterfall Plots for Information processing $I_v(t=0)$ of both PFC and HC regions for year 0, (c) and (d) Force Plots for Information processing $I_v(t=0)$ of both PFC and HC regions for year 0, (e) and (f) Decision Plots for Information processing $I_v(t=0)$ of both PFC and HC regions for year 0

The bar plot in Figure 3a offers a global explanation of the RL model used in this experiment. The mean absolute value for each feature in each class is calculated and plotted. The order of importance for features is as follows: $I_{HC}(t-1)$, $I_{PFC}(t-1)$,X_{HC}, X_{PFC}, D_{HC}, and D_{PFC}. In [20] brain cognition is attributed to brain activity, $Y_v(t)$, and amyloid accumulation, $D_v(t)$. [20] also highlights the direct correlation between $Y_v(t)$, $X_v(t)$ and $I_v(t)$. Hence, By plotting $I_v(t-1)$ and $X_v(t)$ as most important features, Figure 3a illustrates brain activity having greater significance in the RL model's prediction of $\Delta I_v(t)$ than amyloid accumulation.

Similar to the bar plot, the beeswarm plots depicted in Figures 3b and 3c identify features $I_v(t-1)$ and $X_v(t)$ as being most impactful in the model's prediction of $\Delta I_v(t)$ for the two regions. The beeswarm plot contextualizes model behavior even more. Figure 3b shows that low feature values of $I_{PFC}(t-1)$, $I_{HC}(t-1)$ and X_{HC} and increase the predicted $\Delta I_{PFC}(t)$ while high feature values decrease the predicted change. Figure 3c visualizes that low feature values for $I_{PFC}(t-1)$ and $I_{HC}(t-1)$ increase the model's prediction of $\Delta I_{HC}(t)$ while high feature values decreases the model's prediction.

Figure 4 explains how each feature impacts the model's prediction of $\Delta I_v(t)$ for the Prefrontal Cortex and the Hippocampus for a single patient, for a particular year. Figure 4a, 4c, and 4e explain the Prefrontal Cortex while Figure 4b, 4d and 4f explain the Hippocampus region. The observed patterns in global explanation plots are also visible in the plots for local explanations. $I_v(t)$ is the most important feature in predicting $\Delta I_v(t)$. $X_v(t)$ is also an important feature in predicting $\Delta I_v(t)$. These local and global explanations provide evidence for the assertion that brain activity plays a more pivotal role in brain degradation than amyloid accumulation. In [29] and [30], authors discuss the prominent Alzheimer's disease (AD) theory centered on β-amyloid (Aβ) protein deposition initiating cognitive decline. [29] explores the connection between lifelong brain activity patterns and Aβ deposition and suggests that manipulating neural activity could impact Aβ levels.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5: SHAP Plots and Trajectories for Patient RID 2219: (a) RL Prediction vs. Ground Truth for Total Cognition (MMSE score), (b) Summary Bar Plot for the change in Information Processing of PFC region $\Delta I_{PFC}(t)$ (Class 0) and HC region $\Delta I_{HC}(t)$ ($I_{HC}(t=0)$), (c) RL Prediction vs. Ground Truth for PFC size, (d) Stacked Force Plot of PFC size X_{PFC} effect on Mean Total Cognition $C(t)$ per year, (e) RL Prediction vs. Ground Truth for HC size, (f) Stacked Force Plot of HC size X_{HC} effect on Mean Total Cognition $C(t)$ per year.

5.1 Per-Patient Analysis

We also conducted an analysis on each patient to determine the effect these factors have on individuals as the disease progresses. Figure 5 shows the results for Patient RID 2219, who was selected for their maximum decrease in MMSE score. Subfigures 5(a), 5(c), and 5(e) show the accuracy of our model in predicting the total cognition $C(t)$, Prefrontal Cortex Size X_{PFC}, and Hippocampus Size X_{HC} respectively as Alzheimer's disease progresses in the patient. Subfigure 5(b) shows how each feature affects the final SHAP value of the model, which corresponds to the change in cognition $\Delta C(t)$. In this patient, the Prefrontal Cortex region has greater effect than the Hippocampus, and the size of the region X has more effect than the information processing $I(t-1)$ of the region. This may be due to the large error in the RL prediction for the Prefrontal

Cortex size, which the model predicted to decrease far more than it actually did. Subfigures 5(d) and 5(f) show the effect of X_{PFC} and X_{HC} in their respective regions on the change in cognition. For both regions, they initially contribute positively to the change in cognition, but as brain region size decreases, their effect also decreases and becomes negative. Guo et al. [31] observed similar effects, where a larger initial brain size helped slow down Alzheimer's progression, but as degradation proceeded it also increased in rate. From these figures, we can conclude that Alzheimer's progression is dynamic and dependent upon the individual it affects.

6 Conclusion

We demonstrate the use of RL and SHAP to predict critical factors in Alzheimer's Disease progression. Our Reinforcement Learning model forecasts Prefrontal Cortex and Hippocampus information processing, brain size, and amyloid accumulation for up to a decade post-diagnosis. SHAP analysis revealed that increased information processing and reduced brain region size significantly contribute to brain degradation in both areas. Future work should analyze these factor trajectories over time to identify patients with the sharpest declines and determine the most impactful factors using SHAP. Our research aims to aid neurologists and researchers in Alzheimer's causality determination and treatment planning..

References

[1] Zeinab Breijyeh and Rafik Karaman. Comprehensive review on alzheimer's disease: Causes and treatment. *Molecules*, 25(24):5789, 2020.

[2] AP Porsteinsson, RS Isaacson, Sean Knox, MN Sabbagh, and I Rubino. Diagnosis of early alzheimer's disease: clinical practice in 2021. *The journal of prevention of Alzheimer's disease*, 8:371–386, 2021.

[3] Alfredo Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, 32(24): 18069–18083, 2020.

[4] Bojan Bogdanovic, Tome Eftimov, and Monika Simjanoska. In-depth insights into alzheimer's disease by using explainable machine learning approach. *Scientific Reports*, 12(1):6508, 2022.

[5] George A Vouros. Explainable deep reinforcement learning: state of the art and challenges. *ACM Computing Surveys*, 55(5):1–39, 2022.

[6] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

[7] Petra E Vértes, Aaron F Alexander-Bloch, Nitin Gogtay, Jay N Giedd, Judith L Rapoport, and Edward T Bullmore. Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences*, 109(15):5868–5873, 2012.

[8] Wei Li, Miao Wang, Wenzhen Zhu, Yuanyuan Qin, Yue Huang, and Xi Chen. Simulating the evolution of functional brain networks in alzheimer's disease: exploring disease dynamics from the perspective of global activity. *Scientific reports*, 6(1):34156, 2016.

[9] Stefan Frässle, Ekaterina I Lomakina, Lars Kasper, Zina M Manjaly, Alex Leff, Klaas P Pruessmann, Joachim M Buhmann, and Klaas E Stephan. A generative model of whole-brain effective connectivity. *Neuroimage*, 179:505–529, 2018.

[10] Herman Galioulline, Stefan Frässle, Samuel J Harrison, Inês Pereira, Jakob Heinzle, and Klaas Enno Stephan. Predicting future depressive episodes from resting-state fmri with generative embedding. *NeuroImage*, 273:119986, 2023.

[11] Wolfgang Fruehwirt, Adam D Cobb, Martin Mairhofer, Leonard Weydemann, Heinrich Garn, Reinhold Schmidt, Thomas Benke, Peter Dal-Bianco, Gerhard Ransmayr, Markus Waser, et al. Bayesian deep neural networks for low-cost neurophysiological markers of alzheimer's disease severity. *arXiv preprint arXiv:1812.04994*, 2018.

[12] Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sebastien Ourselin, Nick C Fox, et al. An event-based model for disease progression and its application in familial alzheimer's disease and huntington's disease. *NeuroImage*, 60(3):1880–1889, 2012.

[13] Neil P Oxtoby, Alexandra L Young, David M Cash, Tammie LS Benzinger, Anne M Fagan, John C Morris, Randall J Bateman, Nick C Fox, Jonathan M Schott, and Daniel C Alexander. Data-driven models of dominantly-inherited alzheimer's disease progression. *Brain*, 141(5): 1529–1544, 2018.

[14] Lingyu Liu, Shen Sun, Wenjie Kang, Shuicai Wu, and Lan Lin. A review of neuroimaging-based data-driven approach for alzheimer's disease heterogeneity analysis. *Reviews in the Neurosciences*, 2023.

[15] Weiming Lin, Tong Tong, Qinquan Gao, Di Guo, Xiaofeng Du, Yonggui Yang, Gang Guo, Min Xiao, Min Du, Xiaobo Qu, et al. Convolutional neural networks-based mri image analysis for the alzheimer's disease prediction from mild cognitive impairment. *Frontiers in neuroscience*, 12:777, 2018.

[16] Solale Tabarestani, Maryamossadat Aghili, Mehdi Shojaie, Christian Freytes, and Malek Adjouadi. Profile-specific regression model for progression prediction of alzheimer's disease using longitudinal data. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1353–1357. IEEE, 2018.

[17] Krishnakant Saboo, Chang Hu, Yogatheesan Varatharajah, Prashanthi Vemuri, and Ravishankar Iyer. Predicting longitudinal cognitive scores using baseline imaging and clinical variables. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1326–1330. IEEE, 2020.

[18] Tory O Frizzell, Margit Glashutter, Careesa C Liu, An Zeng, Dan Pan, Sujoy Ghosh Hajra, Ryan CN D'Arcy, and Xiaowei Song. Artificial intelligence in brain mri analysis of alzheimer's disease over the past 12 years: A systematic review. *Ageing Research Reviews*, 77:101614, 2022.

[19] Nair Bini Balakrishnan, PS Sreeja, and Jisha Jose Panackal. Alzheimers disease diagnosis using machine learning: A review. *arXiv preprint arXiv:2304.09178*, 2023.

[20] Krishnakant Saboo, Anirudh Choudhary, Yurui Cao, Gregory Worrell, David Jones, and Ravishankar Iyer. Reinforcement learning based disease progression model for alzheimer's disease. *Advances in Neural Information Processing Systems*, 34:20903–20915, 2021.

[21] Erika Puiutta and Eric MSP Veith. Explainable reinforcement learning: A survey. In *International cross-domain conference for machine learning and knowledge extraction*, pages 77–95. Springer, 2020.

[22] Alejandro Barredo Arrieta, Natalia Diaz-Rodriguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[23] Alexandre Heuillet, Fabien Couthouis, and Natalia Díaz-Rodríguez. Explainability in deep reinforcement learning. *Knowledge-Based Systems*, 214:106685, 2021.

[24] Stephanie Milani, Nicholay Topin, Manuela Veloso, and Fei Fang. A survey of explainable reinforcement learning. *arXiv preprint arXiv:2202.08434*, 2022.

[25] Yunpeng Qing, Shunyu Liu, Jie Song, and Mingli Song. A survey on explainable reinforcement learning: Concepts, algorithms, challenges. *arXiv preprint arXiv:2211.06665*, 2022.

[26] Satyam Kumar, Mendhikar Vishal, and Vadlamani Ravi. Explainable reinforcement learning on financial stock trading using shap. *arXiv preprint arXiv:2208.08790*, 2022.

[27] Ali K Raz, Sean Matthew Nolan, Winston Levin, Kshitij Mall, Ahmad Mia, Linas Mockus, Kris Ezra, and Kyle Williams. Test and evaluation of reinforcement learning via robustness testing and explainable ai for high-speed aerospace vehicles. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–14. IEEE, 2022.

[28] Lloyd S. Shapley. A value for n-person games. In *Contributions to the Theory of Games*, volume II, pages 307–317. Princeton University Press, 1953.

[29] William J. Jagust and Elizabeth C. Mormino. Lifespan brain activity, β-amyloid, and alzheimer's disease. *Trends in Cognitive Sciences*, 2011.

[30] Harald Hampel, John Hardy, Kaj Blennow, Christopher Chen, George Perry, Seung Hyun Kim, Victor L Villemagne, Paul Aisen, Michele Vendruscolo, Takeshi Iwatsubo, et al. The amyloid-β pathway in alzheimer's disease. *Molecular psychiatry*, 26(10):5481–5503, 2021.

[31] Liang-Hao Guo, Panagiotis Alexopoulos, Stefan Wagenpfeil, Alexander Kurz, Robert Perneczky, Alzheimer's Disease Neuroimaging Initiative, et al. Brain size and the compensation of alzheimer's disease symptoms: a longitudinal cohort study. *Alzheimer's & Dementia*, 9(5):580–586, 2013.

[32] Garage Contributors. Garage: A toolkit for reproducible reinforcement learning research. *GitHub repository. Opgehaal van https://github. com/rlworkgroup/garage*, 2019.

Supplementary Materials

A The RL Model

Brain Structure: The brain is modeled as a graph $G_s = (V, E)$, where each node $v \in V$ represents a brain region, and each edge $e \in E$ represents a tract. Let $X_v(t)$ indicate the size of brain region $v \in V$ at time t, and let $X(t) = [X_1(t), X_2(t), \ldots, X_{|V|}(t)]$.

Pathology Propagation: The propagation of amyloid, the change of amyloid in a region over time, is modeled below. A network diffusion model is used because it captures the propagation of Aβ through tracts. $D_v(t)$ is the instantaneous amyloid accumulation in region $v \in V$ at time t.

$$\frac{dD_v(t)}{dt} = -\beta H D_v(t) \tag{3}$$

where $D(t) = [D_1(t), D_2(t), \ldots, D_{|V|}(t)]$, H is the Laplacian of the adjacency matrix of the graph G_s, and β is a constant. The total amyloid in a region $\phi_v(t)$:

$$\phi_v(t) = \int_0^t D_v(s)\,ds \tag{4}$$

Brain activity and cognition: To support cognition, multiple brain regions work in synchrony. The activity $Y_v(t)$ in region $vinV$ in support of cognition $C(t)$ at time t. The hypothetical term information processing, $I_v(t) \in R_{\geq 0}$, is introduced to relate a region's size and activity to its "contribution" to cognition. The resulting model for cognition, $C(t)$, supported by the brain at time t can be modeled as:

$$C(t) = \sum_{v \in V} I_v(t) \tag{5}$$

The activity, $Y_v(t)$, in a region depends on both its information processing and its size. The relationship between activity and information processing is proportional, while the relationship between activity and size is inversely proportional. The relationship between the three features is modeled as:

$$Y_v(t) = \gamma \frac{I_v(t)}{X_v(t)} \quad \forall v \in V \tag{6}$$

Energetic cost: To support cognition the brain consumes energy. The energy consumption for a region is proportional to the activity in that region. Therefore the total energy cost of the brain can be modeled as:

$$M(t) = \sum_{v \in V} Y_v(t) \tag{7}$$

Degeneration of brain regions: Neurodegeneration, the change in brain size, is influenced by two factors: amyloid deposition and brain activity. Previous equations and models inferred a linear relationship between the rate at which a brain region degenerates and Aβ deposition. Additionally, brain degeneration can also be accelerated by brain activity. The following equation is a representation of how brain activity, neurodegeneration, and Aβ are related

$$\frac{dX_v(t)}{dt} = -\alpha_v D_v(t) - \alpha_v Y_v(t) \forall v \in V \tag{8}$$

Parameter constants of equations: The demographics of individual patients can affect the progression of AD. To account for the influence of demographics in the model, parameters α_1, α_2, β, γ were used in previous equations. For demographics at baseline Z_0 we define a function f such that

$$(\alpha_1, \alpha_2, \beta, \gamma) = f(Z_0) \tag{9}$$

B Experimental Setup

In order to use the SHAP method to explain model predictions, we first train an RL model on longitudinal data (10-years) using k-fold cross validation. The data input to the model comprises information such as patient ID, diagnosis at baseline year, demographic features (age, gender, education, presence of APOE gene), as well as baseline measurements of brain region size $X(t)$, amyloid accumulation $D(t)$, and information processing of each brain region $I(t)$. The network consists of a stochastic two-hidden-layer Gaussian MLP with 32 neurons per-layer. We experimented with RL algorithms (TRPO and PPO) provided by the Garage library [32]. The training process involves running the model on the train split of each fold (5-fold cross validation) for 1M timesteps (1000 epochs, with a batch size of 1000). Subsequent evaluation involves loading the trained model, running a simulation for each patient (10 year prediction based on baseline year information) and recording observations and actions. The recorded observations are then fed into the SHAP library, where we utilize the model-agnostic Kernel Explainer object to generate global and local (patient-level) explanations of the model's predictions.

C Hyperparameters

We provide the hyperparameters for our experiments in Table 1. These values are used in all our experiments unless specified otherwise.

Hyperparameters	Values
Batch size	1000
Epochs	1000
GAE λ	0.97
Cognition Score	MMSE
Max Cognition (C_{task})	10.0
Scale Observations	True
Action limit	± 2.0
Max timesteps	11 (years)
Number of seeds (per fold)	5

Table 1: Hyperparameters for our experiments with PPO and TRPO

D RL Methods

D.1 TRPO

Trust Region Policy Optimization (TRPO) is an on-policy algorithm with a monotonic improvement guarantee. TRPO is designed to address some of the challenges associated with policy optimization. The key idea to TRPO is to constrain the local variation of the parameters to a "trust region" in the policy-space to ensure the update steps of the policy remains predictive. The constrain on the variation of parameter is determined by KL Divergence. The problem can be described as:

$$\begin{aligned} max \qquad & E_{(s_t,a_t)\sim\pi}[\frac{\pi_\theta(a_t|s_t)}{\pi(a_t|s_t)}\hat{A}_\pi(s_t,a_t)] \\ \text{s.t.} \quad & D_{KL}(\pi_\theta(.|s)||\pi(.|s)) \leq \delta, \forall s \end{aligned} \tag{10}$$

TRPO is known for its stability and ability to handle complex, high-dimensional action spaces, however it can be computationally expensive and requires careful tuning of hyperparameters. Further research on policy optimization methods yielded improvements to TRPO such as Proximal Policy Optimization (PPO).

D.2 PPO

Proximal Policy Optimization (PPO) is a model-free policy gradient algorithm designed to address the limitations of previous policy optimization methods like TRPO. PPO utilizes deep neural networks to learn a policy π_θ and a value function V_ϕ. To train the policy, PPO collects trajectories τ from the old policy network, $\pi_{\theta_{old}}$, prior to updating the network. The policy network is then trained with the trajectories τ to maximize a clipped surrogate policy J_π.

$$J_\pi(\theta) = \mathbb{E}_{s_t, a_t \sim \tau}\left[\min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}\hat{A}_t, \text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}, 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right] \tag{11}$$

Along with the policy network, the value network is also trained with the collected trajectories to minimize the objective J_V.

$$J_V(\phi) = \mathbb{E}_{s_t \sim \tau}\left[(V_\phi(s_t) - \hat{R}_t)^2\right] \tag{12}$$

where $\hat{R}_t = \hat{A}_t + V_\phi(s_t)$ is the value function target. Using the generalized advantage estimator (GAE), the value function target computes the advantage estimates. In practice, the two networks are jointly optimized with shared parameters (i.e., $\theta = \phi$). PPO performs well in both single-agent and multi-agent scenarios and is known to achieve state-of-the-art results in various RL benchmarks.