Discovering a reaction-diffusion model for Alzheimer's disease by combining PINNs with symbolic regression

Zhen Zhang[*], Zongren Zou[*], Ellen Kuhl[†‡], George Em Karniadakis[*‡]

Abstract

Misfolded tau proteins play a critical role in the progression and pathology of Alzheimer's disease. Recent studies suggest that the spatio-temporal pattern of misfolded tau follows a reaction-diffusion type equation. However, the precise mathematical model and parameters that characterize the progression of misfolded protein across the brain remain incompletely understood. Here, we use deep learning and artificial intelligence to discover a mathematical model for the progression of Alzheimer's disease using longitudinal tau positron emission tomography from the Alzheimer's Disease Neuroimaging Initiative database. Specifically, we integrate physics informed neural networks (PINNs) and symbolic regression to discover a reaction-diffusion type partial differential equation for tau protein misfolding and spreading. First, we demonstrate the potential of our model and parameter discovery on synthetic data. Then, we apply our method to discover the best model and parameters to explain tau imaging data from 46 individuals who are likely to develop Alzheimer's disease and 30 healthy controls. Our symbolic regression discovers different misfolding models $f(c)$ for two groups, with a faster misfolding for the Alzheimer's group, $f(c) = 0.23c^3 - 1.34c^2 + 1.11c$, than for the healthy control group, $f(c) = -c^3 + 0.62c^2 + 0.39c$. Our results suggest that PINNs, supplemented by symbolic regression, can discover a reaction-diffusion type model to explain misfolded tau protein concentrations in Alzheimer's disease. We expect our study to be the starting point for a more holistic analysis to provide image-based technologies for early diagnosis, and ideally early treatment of neurodegeneration in Alzheimer's disease and possibly other misfolding-protein based neurodegenerative disorders.

Keywords: Alzheimer's disease, misfolded tau protein, model discovery, uncertainty quantification, PINNs, symbolic regression

1 Introduction

The tau protein plays a critical role in Alzheimer's disease. In healthy brains, tau helps stabilize microtubules to maintain cell structure and transport nutrients. In Alzheimer's disease, tau undergoes abnormal modifications, it misfolds, and forms toxic tangles in the brain. The accumulation of misfolded

[*]Division of Applied Mathematics, Brown University, Providence, RI 02912, USA (zhen_zhang1@brown.edu, zongren_zou@brown.edu, george_karniadakis@brown.edu).

[†]Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA (ekuhl@stanford.edu).

[‡]Corresponding authors.

tau contributes to the spread of pathology throughout the brain and correlates with cognitive decline and neurodegeneration. Understanding the spatio-temporal evolution of tau misfolding is vital for developing interventions that target tau pathology and potentially slow down the progression of Alzheimer's disease.

Until about ten years ago, the only method to diagnose Alzheimer's disease non-invasively in vivo was cognitive testing to confirm memory loss at the very advanced stages of neurodegeneration. Throughout the past decade, the ability to trace tau protein through positron emission tomography has drastically changed how we can image disease progression across the living brain, non-invasively, at any stage. By tracking the accumulation and distribution of misfolded tau over time, these imaging techniques provide insights into disease progression and staging. They allow us to correlated alterations in misfolded tau concentration to cognitive decline, and hold the potential to evaluate therapeutic interventions and monitoring their effectiveness in slowing down disease progression. Towards this goal, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database has become the go-to repository of clinical, imaging, and biomarker data from individuals with normal cognition, mild cognitive impairment, and Alzheimer's disease. It contains dozens of freely available, fully annotated, longitudinal scans and enables researchers to analyze and correlate data.

Computational modeling provides a valuable approach to accurately measure the accumulation and dissemination of two key proteins, misfolded tau and Amyloid-β, that are instrumental in the pathological development of AD. Mathematical models inspired by reaction-diffusion systems hold significant potential for personalized predictions of disease progression timelines. Previous studies have used cross-sectional positron emission tomography (PET) data to calibrate and validate computational models for tau pathology [1, 2]. A simple but efficient model is the Fisher Kolmogorov model combined with network diffusion within the brain's connectome. Mattia et. al. utilized such framework to study the spread of Amyloid-β in the human brain [3]. We have recently shown that this method can be embedded into a hierarchical Bayesian analysis to explain tau protein misfolding and spreading across 83 brain regions from longitudinal neurimaging data of 76 subjects to distinguish amyloid positive patients and healthy controls [4, 5]. However, this method *a priori postulate a mathematical model* of Fisher-Kolmogorov type for the reaction term and does not allow for alternative functional forms to characterize the complex dynamics of protein misfolding.

In our study, we leverage Physics-Informed Neural Networks (PINNs) [6] in conjunction with symbolic regression [7, 8] to unravel the complex nonlinear dynamics of Alzheimer's disease purely from clinical data. This methodological approach mirrors the innovative work of Zapf et al., who employed PINNs to accurately estimate the diffusion coefficient that governs the long-term spread of molecules in the human brain, as evidenced by magnetic resonance images [9]. The increasing adoption of PINNs, as seen in our work and that of others [10, 11, 12], highlights the method's potential in extracting meaningful parameters from real-world data to enhance our understanding of various phenomena. To a priori guarantee physical constraints such as thermodynamic consistency or polyconvexity, recent studies have proposed to hardwire our prior constitutive knowledge into the network input, output, architecture, and activation functions, a strategy that has become known as constitutive artificial neural networks (CANN) [13]. This type of custom-designed neural networks has been used to discover the model, parameters, and experiment that best explain the behavior of human brain [14] and skin [15].

The rest of this paper is organized as follows: in Section 2, we present the problem formulation and the methodology developed for model discovery (see the PINN method in Section 2.1 and the symbolic regression method in Section 2.2); in Section 3, details of the experimental setup and results are discussed, where our approach is first tested on simulated data in Section 3.1 and then used to discover reaction models for amyloid positive and negative patients from real data in Section 3.2; a discussion regarding the proposed methodology and results is provided in Section 4.

2 Methodology

In this section, we formulate the problem and describe the machine learning methodology to discover models for Alzheimer's disease. We focus on the discovery of the reaction model in the reaction-diffusion system [16], described by the following partial differential equation (PDE):

$$\frac{\partial c}{\partial t} = \nabla \cdot (\mathbf{D} \cdot \nabla c) + f(c), \tag{1}$$

where c is the concentration of misfolded tau protein, $\mathbf{D}$ denotes the heteroscedastic diffusion tensor, and $f : \mathbb{R} \to \mathbb{R}$ is the reaction model. Different reaction models result in various behaviors of the system, e.g. [17, 18, 19, 20] (see also Table 1), and the selection of the appropriate model is governed by the nature of the underlying engineering or life sciences problem.

Physics-informed neural networks (PINNs) have proven to be powerful tools to solve differential equations as well as infer unknown quantities from data and physical law; see [21] for a review and [6, 22, 23, 24, 25, 26, 27, 28, 29] for recent developments of PINNs. In this work, PINNs are employed to learn $f(c)$ from data of c and the physics defined in Eq. (1), followed by symbolic regression to determine the analytic expression of f. A similar approach was proposed in [30], in which the PINN and AI Feynman [31] methods were used to discover differential operators in equations from data. Instead, in this paper, we focus on the topic of model discovery.

We follow the same setup as in the literature [32, 4, 33] and model the aggregation and propagation of pathological tau in the connectome of the brain by the reaction-diffusion system defined by Eq. (1) and discretized on a weighted, undirected graph G [34, 35].The nodes of G represent different non-overlapping brain regions and the edges of G are the axonal connections between different regions. In this regard, we discretize Eq. (1) on G, and we recast the PDE as an ODE system:

$$\frac{dc_i}{dt} = h^i_\kappa(t, \mathbf{c}) + f(c_i), i = 1, ..., N, \tag{2}$$

where $\mathbf{c} := [c_1, ..., c_N]^T \in \mathbb{R}^N$ represents the concentration of misfolded tau protein in the N different brain regions (see Figs. 4 and 6 for the illustration and Sec. 3 for the description of selected brain regions), $h^i_\kappa : \mathbb{R}^{N+1} \to \mathbb{R}$ the diffusion term between different regions, and $f : \mathbb{R} \to \mathbb{R}$ the local reaction/production model, which characterizes the collective dynamics of protein production, clearance, and conversion from healthy to unhealthy seeds [36]. More details and reasoning of this setup can be found in [32, 4, 33] and will be discussed in Section 3. Here the diffusion model, h_κ, is assumed to be known and parameterized by the diffusion coefficient κ, while the reaction/production model, f, will be discovered. Our approach for discovering f is formulated as follows: first infer $\mathbf{c}$, f, and κ from data of c and the physics defined in Eq. 2 using PINNs, and then find the analytic expression of f using symbolic regression. We remark that in this approach, we use two individual NNs: one for the approximation of $\mathbf{c}$, which takes the time t as input, and one for the approximation of f, which takes the misfolded protein concentration c_i as input. Importantly, the data of $\mathbf{c}$, generated either synthetically from simulations or collected longitudinally from medical images, are only used in the first step. A schematic view of the workflow is illustrated in Fig. 1, assuming a homogeneous distribution with $N = 1$ and no diffusion, and provided in Procedure 1. A pedagogical example on discovering the Kraichanan-Orszag dynamical system [37, 10, 38] and a tutorial of this approach can be found in Appendix A for better understanding.

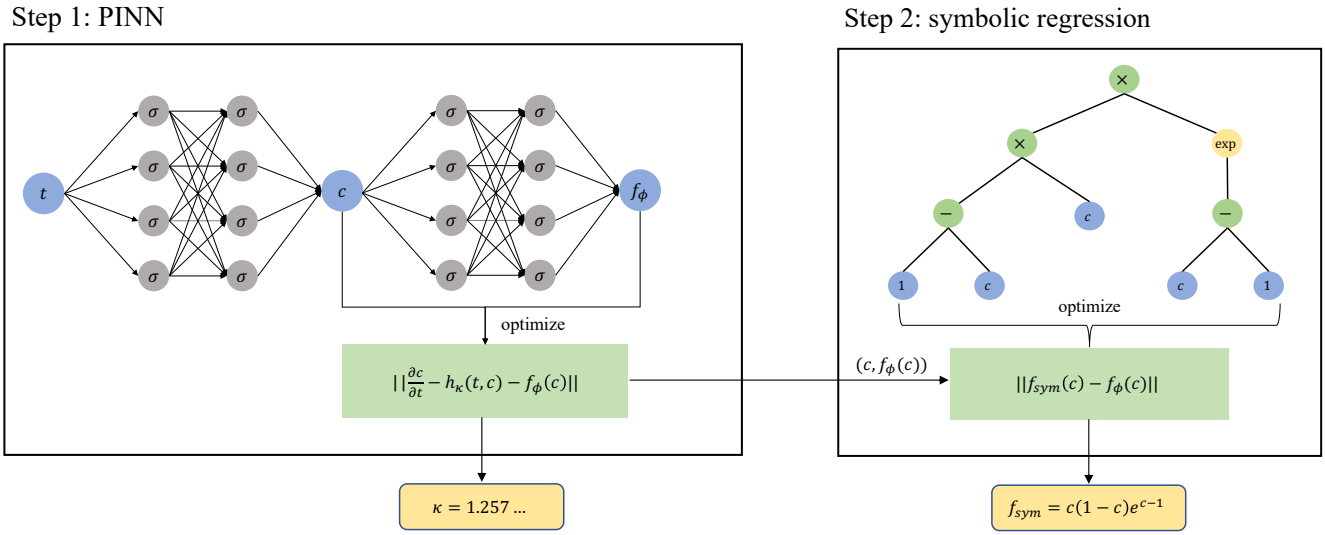

Figure 1: **A schematic view of the approach for model discovery**. The original PINN framework for inferring the diffusion coefficient κ and the reaction model f (**left**) is followed by a symbolic regression step to discover the analytic form of f (**right**).

Procedure 1 Identify κ and the analytic form of f in Eq. (2)

Step 0 Specify two individual NNs, denoted as $\mathbf{c}_\theta$ and f_ϕ: $\mathbf{c}_\theta$ approximates the solution to Eq. (2), which takes as input t and outputs $\mathbf{c}$, while f_ϕ approximates the local production model, which takes as input $c_i, i = 1, ..., N$ and outputs the local production term for the i^{th} region.

Step 1 Follow the PINN framework [6] and minimize the loss function defined in (3), such that $\mathbf{c}_\theta$ fits the data and $\mathbf{c}_\theta$, κ, and f_ϕ satisfy the equation (2).

Step 2 Regress the analytic form of f using a symbolic regression method [7] from data of c_i and $f_\phi(c_i)$.

2.1 Physics-informed neural networks (PINNs)

The PINN method, originally proposed by Raissi et. al. in [6], addresses ODE/PDE problems by deploying neural networks (NNs) as surrogate models for quantities of interest, constructing a loss function with respect to the ODE/PDE used to describe the physics, and optimizing NN parameters such that the loss function is minimized (see [6, 22] for more details). In our case, $\mathbf{c}$ and f are approximated with NNs, denoted by $\mathbf{c}_\theta$ and f_ϕ, respectively, where θ and ϕ denote NN parameters, e.g., weights and biases. Let $\mathcal{T}_D$ denote the set of t on which data of $\mathbf{c}$ is (partially) available and $\mathcal{T}_R$ denote the set of t on which residuals of the equation are computed. Gradient-descent based methods, e.g., Adam [39] and L-BFGS [40], can be employed to minimize the following loss function:

$$\mathcal{L}(\Theta) = \mathcal{L}_{data}(\theta) + \mathcal{L}_{res}(\theta, \phi, \kappa) + \mathcal{L}_{aux}(\theta, \phi, \kappa) \tag{3}$$

with respect to $\Theta := \{\theta, \phi, \kappa\}$, where

$$\mathcal{L}_{data}(\theta) = \frac{\sum_{t \in \mathcal{T}_D} ||\mathbf{c}(t) - \mathbf{c}_\theta(t)||_2^2}{|\mathcal{T}_D|}, \tag{4}$$

$$\mathcal{L}_{res}(\theta, \phi, \kappa) = \frac{\sum_{t \in \mathcal{T}_R} ||\frac{d\mathbf{c}_\theta}{dt}(t) - h_\kappa(\mathbf{c}_\theta) - f_\phi(\mathbf{c}_\theta)||_2^2}{|\mathcal{T}_R|}, \tag{5}$$

and $\mathcal{L}_{aux}$ is the auxiliary loss that can be seen as additional physical information. More details will be discussed in Section 3 and in Eq. (12).

It is possible to enforce boundary conditions on f or $\mathbf{c}$ [22, 41, 42]. For example, for f one can set

$$f_\phi(\mathbf{c}) = G(\mathbf{c}) + D(\mathbf{c}) \odot g_\phi(\mathbf{c}), \tag{6}$$

where $D(\mathbf{c})$ represents the distance from $\mathbf{c}$ to the boundary, G is a function that matches f on the boundary, while g_ϕ is a fully-connected neural network.

2.2 Symbolic regression

For symbolic regression, we adopt the open-source software PySR [7], which has a configurable Python interface built on the efficient Julia backend *SymbolicRegression.jl*. The underlying algorithm for PySR involves tree search and regularized evolution. In our case of study, we use PySR to distill knowledge from f_ϕ to obtain a symbolic expression f_{sym}.

Evaluation Metric

We use *score* defined in [8] as one of our evaluation metrics for the correctness of an expression,

$$score = -\Delta \log(\text{MAE})/\Delta C,$$

where MAE is the mean absolute error between the prediction and the data, C refers to the complexity of the expression, and Δ denotes local change [8]. Higher *score* means that with a slightly lower complexity, MAE of the symbolic regression model becomes much larger. A model with low loss and high score is preferred, so we pick the expression with the highest score among those whose loss is lower than 1.5 times the loss of the most accurate model as our final candidate.

When multiple simulations are conducted in the PINN inference stage, one may obtain multiple functions f_ϕ. The score metric can be used to pick one f_{sym} for each f_ϕ. Then, in order to compare f_{sym} obtained from different f_ϕ, we use the projection error as another evaluation metric. We first solve Eq. (2) with the inferred parameter κ from the PINN in step 1 and the discovered model f_{sym} from the symbolic regression in step 2 to obtain the solution, denoted by $\mathbf{c}_{proj}$, and then we evaluate $\mathbf{c}_{proj}$ on $t \in \mathcal{T}_D$. Specifically, the projection error is defined by

$$\text{projection error} = \frac{\sum_{t \in \mathcal{T}_D} ||\mathbf{c}(t) - \mathbf{c}_{proj}(t)||_2^2}{|\mathcal{T}_D|}. \tag{7}$$

The candidate with the lowest projection error is the one that fits the data best.

3 Results

In this section, we discuss details of the experimental setup and discovery of the reaction model of the reaction-diffusion system. We first test the approach with synthetic data in Section 3.1, in which we use data from simulations of Eq. (2) with four different reaction models. Then we move to real data in Section 3.2, where two models are discovered using our approach with uncertainty. Specifically for real data, we assume that all individuals within each group, amyloid positive and amyloid negative, share the

same type of reaction model but personalized with different reaction rates. A deep ensemble method for PINNs [43, 10, 44] is employed for uncertainty quantification.

As mentioned in Section 2, to apply the reaction-diffusion equation to our brain-related problems, we follow the same setup as in [4] and discretize Eq. (1) on a graph G, which comes from the Budapest Reference Connectome v3.0 [45] and the Human Connectome Project [33]; see [4] for details on how to construct G. There are in total $N = 83$ nodes in the graph G, representing 83 considered cortical and subcortical brain regions. In this work, we adopt the graph G and discretization from [4] and Eq. (1) is discretized as follows:

$$\frac{dc_i}{dt} = -\kappa \sum_{j=1}^{N} L_{ij} c_j + \alpha f(c_i), i = 1, ..., N, \tag{8}$$

where c_i is the concentration of tau protein in brain region i, κ determines the transport rate of misfolded protein between regions, (L_{ij}) is the graph Laplacian, representing the connectivity of the graph, and α denotes the reaction rate; κ and α are assumed to be specific to each individual but shared across all regions. We note that this corresponds to Eq. (2) with $h^i_\kappa(\mathbf{c}) = -\kappa \sum_{j=1}^{N} L_{ij} c_j$.

Unlike previous studies [36, 4], which assume a single specific reaction model, $f(c) = c(1 - c)$, herein we aim to identify the analytical form of the function f from tau concentration data of 76 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

The intuition behind this is as follows: Assuming the same reaction model for all amyloid positive and negative subjects might be too constrained and have a negative effect when the fitting of data. However, we found empirically that requirements from the Fisher equation [17] and the Kolmogorov-Petrovsky-Piskunov (KPP) equation [16] on the reaction term were necessary to provide better and more realistic results. As a generalized form of the Fisher equation, the KPP equation requires that the reaction term f has the following properties:

$$f(0) = f(1) = 0, f(c) > 0, \tag{9a}$$

$$f'(c) < f'(0), \forall c \in [0, 1], \tag{9b}$$

which can be considered as the prior knowledge or constraints of the to-be-discovered model. Eq. (9a) implies no creation or depletion of misfolded tau protein when $c = 0$ (no tau protein) or $c = 1$ (the region is completely occupied by tau protein) and the growth rate of tau protein is always positive. Eq. (9b) implies that the reaction process is less intense or "slower" as c increases from 0 to 1. In other words, the reaction process is more significant when tau concentration is close to 0 and becomes less important as c approaches 1. To satisfy constraints in Eqs. (9a) and (9b), we first set

$$\tilde{f}_\phi(c) = c(1 - c)e^{g_\phi(c)} \tag{10}$$

to enforce (9a), where g_ϕ is a standard fully-connected NN parameterized by ϕ. Then $\tilde{f}_\phi$ is normalized by setting

$$f_\phi(c) = \frac{\tilde{f}_\phi(c)}{4 \max_{x \in [0,1]} \tilde{f}_\phi(x)} \tag{11}$$

as our final parameterization of f to ensure the inferred reaction term does not degenerate to 0. We introduce

$$\mathcal{L}_{aux} = \frac{\sum_{c \in \mathcal{C}} || \max(0, \alpha f'_\phi(c) - \alpha f'_\phi(0))||_1}{|\mathcal{C}|}, \tag{12}$$

in (3) as the auxiliary loss term to softly enforce condition (9b), which guarantees a travelling wave solution to the Fisher-Kolmogorov equation. In this paper we set $\mathcal{C} = \{0, 0.01, 0.02, \cdots, 1\}$. Another reason for introducing this additional regularization term is to promote smoothness of the inferred reaction term.

Additional types of KPP equations, including the Newell-Whitehead-Segel equation and the Zeldovich-Frank-Kamenetskii equation, are summarized in Table 1.

Group	Equation	General form	Reference	Our case
1	Fisher	$kc(1-c)$	[17]	$c(1-c)$
2	Newell-Whitehead-Segel	$kc(1-c^q)$	[18, 19]	$\frac{3\sqrt{3}}{8}c(1-c^2)$
3				$\frac{2^{2/3}}{3}c(1-c^3)$
4	Zeldovich-Frank-Kamenetskii	$kc(1-c)e^{\beta(c-1)}$	[20]	$\frac{\sqrt{5}+2}{4}c(1-c)e^{c-1-\frac{\sqrt{5}-3}{2}}$

Table 1: **Reaction term f in KPP equations.** We stratify the subjects into 4 groups with the assumption that subjects in each group share the same reaction term $f(c)$ in Section 3.1. Diverse reaction terms are associated with distinct KPP equations, which are in turn indicative of particular biological characteristics. We normalize f to make its maximum value $\frac{1}{4}$ in all groups.

3.1 Synthetic data

Data preparation

In this study, we conducted a simulation of tau concentration for a sample of 76 subjects with varying initial conditions, parameters, and reaction terms, employing Equation (8). The subjects were stratified into four groups of equal size, with differing reaction models across groups. As detailed in Table 1, the reaction models are chosen from $c(1-c)$ (the spreading of biological populations [17]), $c(1-c^2)$ (a convection in fluid thermodynamics [18, 19]), and $c(1-c)e^{\beta(c-1)}$ (a flame propagation in combustion [20]). We note that these reaction models are chosen for the purpose of verifying our approach with simulated data. The parameters κ, α_i, and α_{ij} were assumed to follow probability distributions of $\text{BoundNormal}(1, 0.5^2)$, $\mathcal{N}(0.6, 0.1^2)$, and $\mathcal{N}(\alpha_i, 0.2^2)$, respectively, where α_{ij} denotes the j-th subject in the i-th group. The initial tau concentration for the i-th node, denoted by $c_i(0)$, was sampled from a normal distribution with equivalent mean and variance as the real data. The connectivity matrix, denoted by L, was set to be identical to the real data. The tau concentrations $c(t_k)$ are sampled at $t_k = k$ for $k = 0, 1, 2$ years.

Parameter and function identification

The initial step in the present study involves the determination of undetermined parameters and remainder terms, specifically κ and $\alpha f(c)$, through the application of PINNs. We set α and κ as subject-specific learnable parameters and approximate $f(c)$ with a neural network f_ϕ, which is hard constrained as indicated in (10). The training loss is shown in the left portion of Figure 3. The outcomes of the identification process are displayed in Figure 2. The left portion of the figure displays the inferred and actual distributions of the transport rate, κ, at the global level, utilizing the Gaussian kernel density estimator [46]. The congruence between the two distributions signifies that κ can be accurately identified. In the middle of the figure, the inferred and actual distributions of the local production rate, κ, at the group level are illustrated. Notably, the PINNs method successfully captures the appropriate distribution of the growth rate α in each group, regardless of the corresponding reaction term. The right portion of the figure presents the inferred and actual reaction terms for each group, demonstrating that the neural network inference result, f_ϕ, concurs with the ground truth f for all four cases. The effect of hard constraining f_ϕ is discussed in Appendix B, where we found that without the hard constraint, f_ϕ eventually deviates from the ground truth on regions where there is no data for c, which affects the

prediction capability of the entire framework. This is illustrated in Figure 7 that we need data up to $t = 6$, meaning medical images for consecutive six years, if there is no constraint on the boundary values.

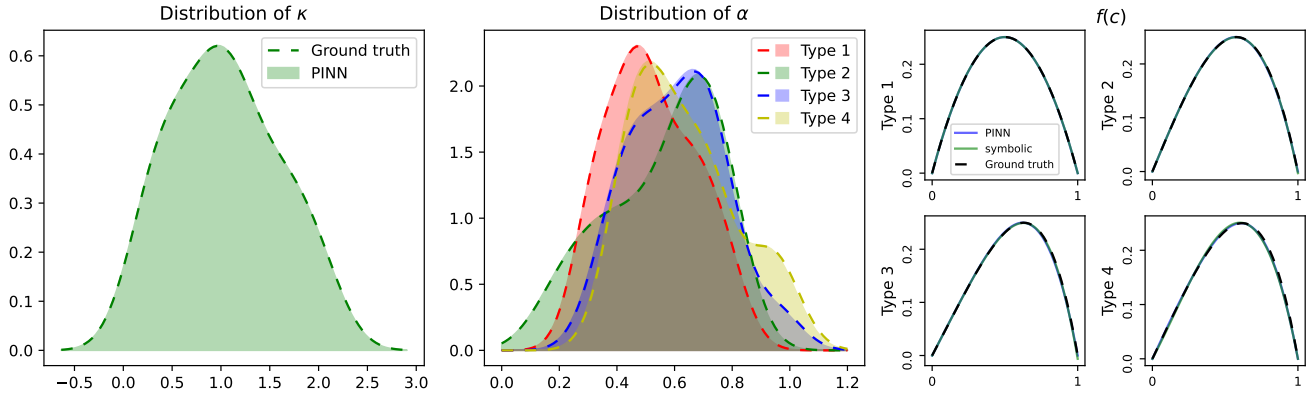

Figure 2: **Parameter and function identification results on synthetic data.** Inference results for transport rate (**left**) and local production rate (**middle**), as well as the reaction term for each group (**right**). PINNs accurately identify the distribution of the transport rate at the global level and the distribution of α in each group. The neural network inference result, f_ϕ, as well as the symbolic regression approximation, f_{sym} agree with the ground truth f for all four cases.

In the second stage of our analysis, we utilize the symbolic regression package PySR to discover closed-form expressions for the four functions f_ϕ. To achieve this, we minimize the mean squared error (MSE) as the target function. Each f_ϕ undergoes 100 iterations of simulation. The binary operators are chosen to be addition, subtraction, and multiplication, the unary operators are exponential and reciprocal, and the complexities of unary operators are set to be 3 and 3. PySR generates analytical expressions, evaluates their corresponding losses (MSE), and scores at every complexity level. This information is presented in the right portion of Figure 3. Subsequently, for each subject, we follow the procedures outlined in Section 2.2 to select a final candidate expression f_{sym} as the associated reaction term. We present the four chosen expressions along with their respective scores highlighted with a black circle. Notably, the discovered f_{sym} aligns with the correct reaction function. Furthermore, a plateau in loss is observed after selecting the best candidate, suggesting that increasing the complexity of the discovered function beyond this point would only result in a minimal improvement of the loss.

Projection of tau concentration

In the last stage of our analysis, after we have identified the unknown parameters κ and α and the unknown functions (f_ϕ and f_{sym}), we substitute them back into the ODE (8) and solve the equation up to $t = 20$ with the built-in Python method *scipy.integrate.odeint* to examine the predictability of our models. The projections of the tau concentration over 20 years after the first PET scan in 36 subjects (nine from each group) and three different brain regions, namely entorhinal cortex (EC), middle temporal gyrus (MTG), and superior temporal gyrus (STG), are illustrated in Fig. 4. We selected these regions due to their significance in Alzheimer's disease (AD). The entorhinal cortex (EC) is typically the first to show AD-related protein misfolding, with changes occurring before symptoms appear. The middle and superior temporal gyrus (MTG and STG) are crucial for short-term memory, which is often impacted early in AD. Studying these areas will help us understand AD's early changes and related cognitive deficits.

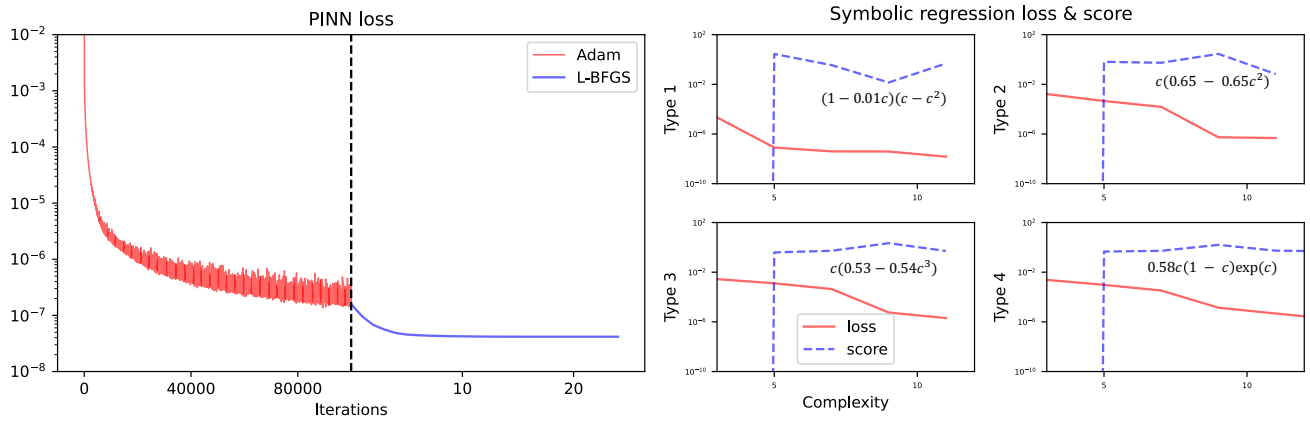

Figure 3: **Losses and evaluation metrics for PINNs and symbolic regression on synthetic data. (Left)** Training loss $\mathcal{L} = \mathcal{L}_{data} + \mathcal{L}_{res}$ for PINN. L-BFGS is applied following Adam to ensure convergence. **(right)** Loss and score for the symbolic regression model at each complexity level. The inferred f_{sym} is highlighted. It aligns with the correct reaction function, and the plateau in loss suggests minimal loss improvement by increasing the expression complexity.

The projection outcome achieved by PINNs, alongside the symbolic regression model, demonstrates a high degree of concordance with the ground truth solution. Moreover, we present the deduced values of α and κ for the 36 subjects, which are located at the top of every subfigure. These inferred values are also in agreement with the preset parameters.

3.2 Real data

Data preparation

We utilize preprocessed tau concentration data directly from the literature [4]. In total 76 subjects are classified into two groups: 46 are identified as amyloid positive ($A\beta^+$) meaning their mean amyloid concentration exceeds a certain level, and 30 are identified as amyloid negative ($A\beta^-$). For each subject, three data points are provided, which are on average one year (1.07 ± 0.31) apart.

Parameter and function identification with uncertainty quantification

Observational noise in the real-life dataset could significantly increase the training difficulty of neural network models as well as complicate the optimization landscape. For such highly nonconvex optimization problem, the parameter initialization becomes more important. Different neural network initialization could lead to drastically different solutions and randomness in the learned parameter set. Even though sometimes it is difficult for the model to identify global minima for every initialization, most modern optimization techniques can handle local minima and ultimately find a parameter set close to the the global minimum [47, 48]. Such uncertainty quantification method is commonly used in machine learning due to its simplicity yet effectiveness [43, 10, 44] and often referred to as deep ensemble. The randomness in the trained NNs in this way can be interpreted as *model uncertainty* [29]. In this work, we conduct ten independent simulations in the PINN inference stage to obtain ten different parameter sets of α and κ for each subject, as well as ten different functions f_ϕ^+ and f_ϕ^- for the amyloid positive and negative groups. The distribution of the personalized parameters κ and α is shown in the left two columns of Fig. 5. Note that we only present the distribution for the best of all ten simulations, i.e., the simulation with lowest

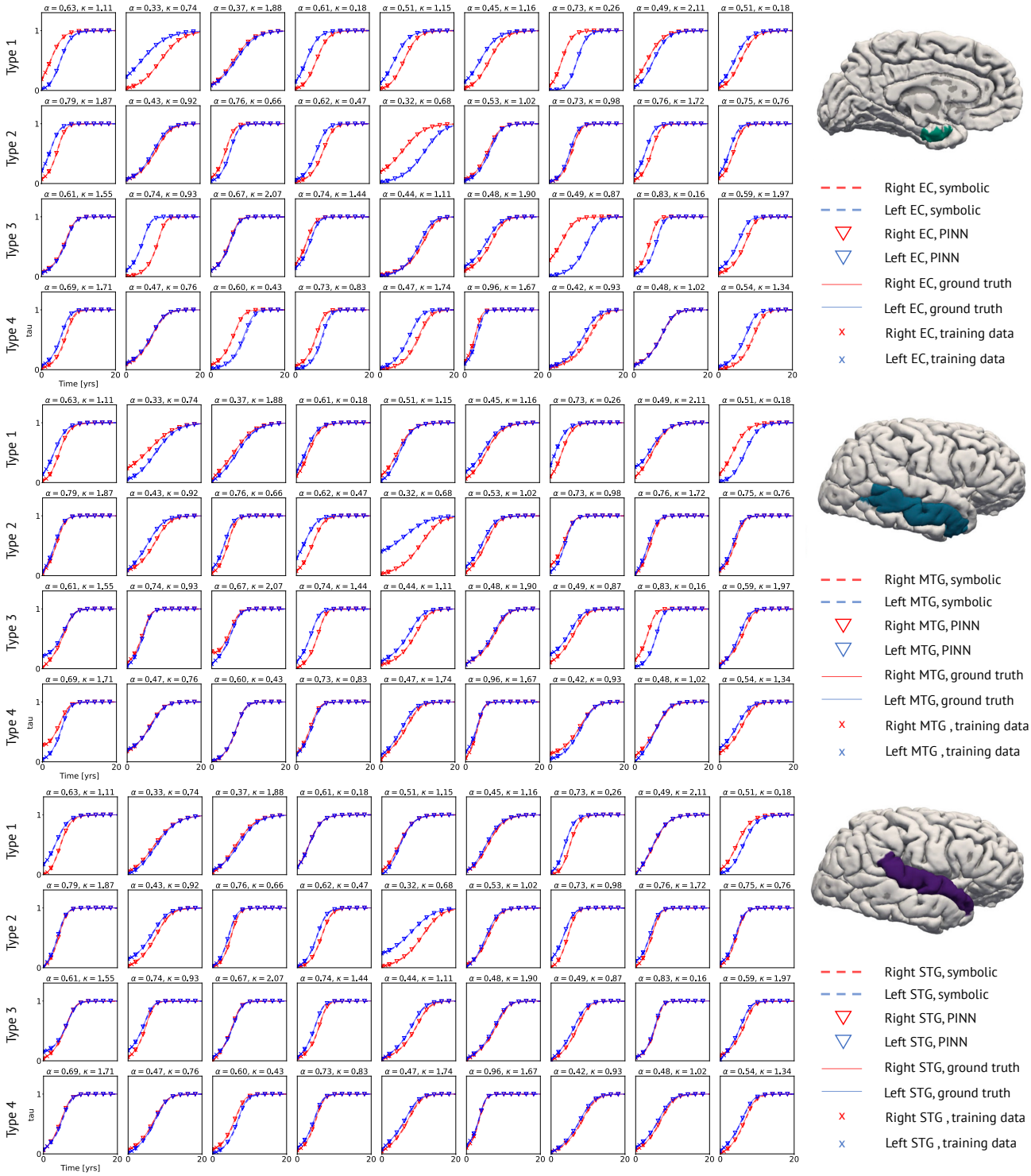

Figure 4: **Prediction of tau concentration.** Evolution of misfolded tau protein in entorhinal cortex (EC), middle temporal gyrus (MTG), and superior temporal gyrus (STG) throughout 20 years based on synthetic data from the first two years.

projection error, since it is not appropriate to aggregate parameter distributions corresponding to different f. On the rightmost column of Fig. 5, we show the ensemble of ten f_ϕ^-, f_ϕ^+, as well as f_{sym}^-, f_{sym}^+. It can be seen that in ten independent simulations, the reaction term clustered into two modes depending on whether the subject is from the amyloid positive or negative group. Interestingly, the variation of the discovered function f_{sym} appears to be larger for the negative group.

Table 2 summarizes the discovered reaction terms for each simulation together with their projection errors, as defined in Section 2.2. For each group, we highlight the expression with lowest projection error. Stricingly, our symbolic regression model discovers different misfolding models $f(c)$ for the two groups, with a steeper increase for the Alzheimer's group, $f(c) = 0.23c^3 - 1.34c^2 + 1.11c$, than for the healthy control group, $f(c) = -c^3 + 0.62c^2 + 0.39c$. We note that although the misfolding model can be deterministically chosen based on the lowest projection error, we present results from all experiments to quantify the model uncertainty for the purpose of reliable and trustworthy machine learning. The error and the uncertainty can be decreased using transfer learning techniques, if an informative prior of the brain model is accessible.

	κ	α	f_{sym}	Projection error
$A\beta^-$	0.49 ± 0.51	0.24 ± 0.41	$-0.64c^3 + 0.64c$	7.19E-04
	0.50 ± 0.57	0.30 ± 0.48	$-c^3 + 0.62c^2 + 0.39c$	4.94E-04
	0.52 ± 0.62	0.30 ± 0.47	$-0.86c^2 + 0.86c$	2.09E-03
	0.60 ± 1.02	0.28 ± 0.45	$-c^3 + 0.59c^2 + 0.41c$	4.95E-04
	0.49 ± 0.48	0.29 ± 0.28	$-0.65c^2 + 0.65c$	6.45E-04
	0.49 ± 0.50	0.24 ± 0.40	$-0.66c^2 + 0.66c$	6.00E-04
	0.51 ± 0.59	0.24 ± 0.40	$-0.64c^2 + 0.64c$	7.08E-04
	0.57 ± 0.88	0.29 ± 0.45	$-c^3 + 0.62c^2 + 0.39c$	**4.80E-04**
	0.51 ± 0.62	0.27 ± 0.45	$-c^3 + 0.57c^2 + 0.43c$	5.01E-04
	0.52 ± 0.64	0.22 ± 0.37	$-0.65c^3 + 0.65c$	6.14E-04
$A\beta^+$	0.30 ± 0.29	0.16 ± 0.16	$-c^2 + c$	1.69E-03
	0.30 ± 0.32	0.18 ± 0.17	$-c^2 + c$	1.74E-03
	0.29 ± 0.28	0.18 ± 0.17	$-c^2 + c$	1.72E-03
	0.31 ± 0.33	0.16 ± 0.16	$-c^2 + c$	1.70E-03
	0.29 ± 0.28	0.15 ± 0.15	$-c^2 + c$	1.70E-03
	0.31 ± 0.33	0.15 ± 0.15	$-c^2 + c$	1.70E-03
	0.30 ± 0.33	0.17 ± 0.16	$0.11c^3 - 1.15c^2 + 1.03c$	1.71E-03
	0.29 ± 0.28	0.16 ± 0.16	$0.07c^3 - 1.08c^2 + 1.02c$	1.71E-03
	0.29 ± 0.28	0.16 ± 0.16	$-c^2 + c$	1.69E-03
	0.29 ± 0.28	0.15 ± 0.14	$0.23c^3 - 1.34c^2 + 1.11c$	**1.68E-03**

Table 2: **Identified α, κ, f_{sym} for the negative and positive groups in 10 independent simulations.** The lowest projection errors for each group are highlighted in bold.

Projection of tau concentration with uncertainty quantification

We substitute the parameters inferred by PINN and reaction terms inferred from the symbolic regression model back to ODE of Eq.(8) and extrapolate in time up to $t = 30$ for 36 subjects and three brain regions, similar to the synthetic data. Since ten independent simulations are conducted, we can propagate the uncertainty in the parameters (α, κ) and functions (f_{sym}^-, f_{sym}^+) to quantify the uncertainty in the

solution c. We obtain the extrapolated tau concentrations in each simulation, and plot their minimum and maximum values at each time step, as shown in Fig. 6. The predicted concentrations given by the model with lowest projection error are plotted in dashed lines.

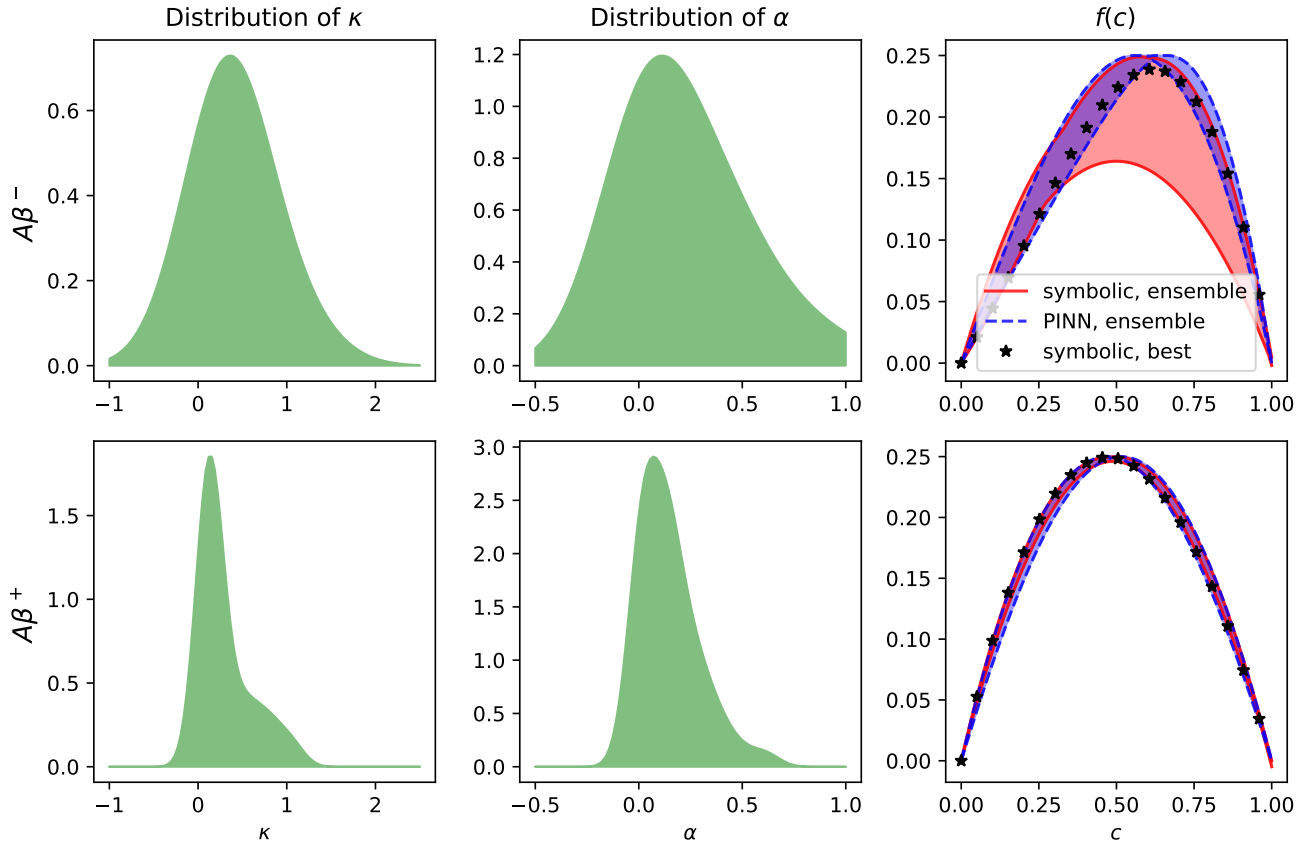

Figure 5: **Parameter and function identification results on real data.** We present the inference results for the transport rate κ and local production rate α, alongside the reaction term for both the positive and negative groups. (**Left, Middle**) To present the obtained results, we only visualized the (population-level) distribution of α and κ in the simulation with the lowest projection error, as it is not appropriate to aggregate the parameter distribution corresponding to different fs. (**Right**) We plot the ensemble of inferred f_{sym} and f_{ϕ}, with shaded region representing the minimum and maximum value of fs in all 10 simulations. We highlight the f_{sym} which corresponds to the simulation with the lowest projection error with stars.

4 Summary and discussion

Tau protein is a key player in Alzheimer's disease, and misfolded tau proteins play a crucial role in disease progression and pathology. Protein misfolding and spreading across the brain of Alzheimer's patients follows a characteristic stereotypical pattern that we can model with reaction-diffusion type equations. There is little controversy about the *diffusion term* of these equation: Since tau is an intracellular protein, the underlying assumption is that it spreads along axons within the brain's connectome. However, the precise nature of the *reaction term* of these equations remains incompletely understood.

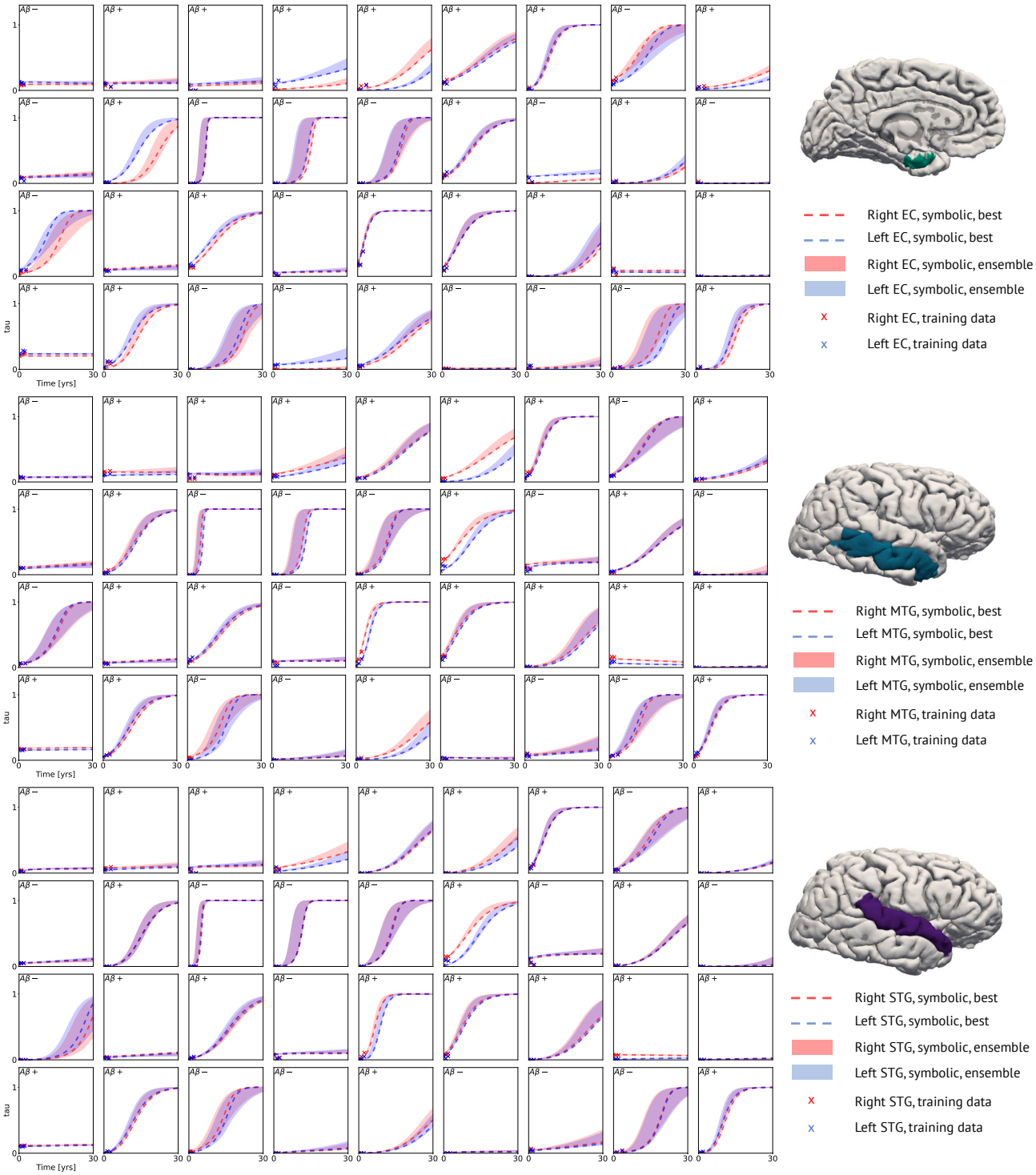

Figure 6: **Projection of tau concentration in entorhinal cortex (EC), middle temporal gyrus (MTG) and superior temporal gyrus (STG) in 30 years based on real data in first three data points.** We plot the ensemble of predicted c given by substituting f_{sym} back to ODE of Eq. (8). The lower and upper bound of the shaded region represents the minimum and maximum value of c in all 10 simulations. We plot the predicted concentrations given by the model with lowest projection error in dashed lines.

Here, we capitalize on the recent developments in physics-informed deep learning and artificial intelligence to discover a mathematical model for the progression of Alzheimer's disease from clinical data. Specifically, we use longitudinal tau positron emission tomography images from 46 individuals who are likely to develop Alzheimer's disease and from 30 healthy controls. Their brain scans are publicly available through the Alzheimer's Disease Neuroimaging Initiative database. By their very nature, positron emission tomography images have a high spatial resolution, while their temporal resolution is low and limited to only a few points in time. In essence, this temporal information is too sparse to infer protein misfolding dynamics from the medical images alone. This motivates the use of a physics-informed approach.

Physics-informed neural networks or PINNs are a type of machine learning model that integrates our prior physics-based knowledge into the training process of neural networks. By incorporating the reaction-diffusion equation of protein misfolding into the loss function, PINNs satisfy the underlying physics by design, and improve the network's accuracy and generalization capabilities. PINNs are particularly powerful for solving complex problems, when the underlying training data are sparse. While they seem well suited to simulate misfolded protein spreading across the brain from sparse clinical data, they will provide no insight into the functional form of the reaction dynamics.

Here, instead of using a stand-alone PINN simulation, we integrate *physics informed neural networks* and *symbolic regression* to discover a reaction-diffusion type partial differential equation for tau protein misfolding and spreading. Importantly, this is a two-step process in which the first step uses the PINN to learn a noninterpretable model from data and the second step uses symbolic regression to *discover the best model and parameters* to explain the reaction term inferred by the PINN. Importantly, our library of possible models contains a variety of popular engineering reaction terms and our parameters are interpretable by design. The proposed approach is different from previous works studying tau pathology in brain in the sense that we learn the reaction term $f(c)$ instead of fixing it as a quadratic term. This flexibility means we make fewer assumptions about the governing equation, which can result in a better fit to the data, when the underlying equation could be misspecified. However, a potential drawback of our "gray-box" modeling approach is that it may lead to higher computational cost.

We demonstrate the features our two-step approach in terms of synthetic and real data and discover the best model and parameters to explain tau imaging data from 46 individuals who are likely to develop Alzheimer's disease and 30 healthy controls. Strikingly, our method discovers different misfolding models for the two groups, with a faster protein misfolding in the Alzheimer's group, $f(c) = 0.23c^3 - 1.34c^2 + 1.11c$, than in the healthy control group, $f(c) = -c^3 + 0.62c^2 + 0.39c$. We anticipate that our two-step modeling strategy generalizes well to other types of partial differential equations with various engineering applications.

Taken together, our results suggest that PINNs, supplemented by symbolic regression, can discover a reaction-diffusion type model to explain misfolded tau protein concentrations in Alzheimer's disease. Understanding the dynamics of tau protein misfolding and its propagation can provide insights into the mechanisms underlying Alzheimer's disease and potentially lead to the development of effective therapeutic interventions. We expect this study to be the starting point for a more comprehensive analysis to provide image-based technologies for early diagnosis, and ideally early treatment, of neurodegeneration in Alzheimer's disease and possibly other misfolding-protein based neurodegenerative disorders.

Appendix

A Pedagogical example: Discovering the Kraichnan-Orszag system

For better understanding of the proposed integration of physics-informed neural networks (PINNs) and symbolic regression, in this section, we provide a pedagogical example that discovers the Kraichnan-Orszag system [37, 10, 38] from data. The system is described by the following ODE system:

$$\begin{aligned} \frac{du_1}{dt} &= e^{-t/10} u_2 u_3, \\ \frac{du_2}{dt} &= u_1 u_3, \\ \frac{du_3}{dt} &= -2 u_1 u_2, \end{aligned} \tag{13}$$

with initial conditions that are typically drawn from a Gaussian distribution. Instead, here we fix the initial conditions as $u_1(0) = 1, u_2(0) = 0.8, u_3(0) = 0.5$, and assume partial knowledge of the dynamics. Specifically, we assume we know the right-hand side of the third equation to be a linear transformation of $u_1 u_2$, and have zero information of the first two equations. Then we can rewrite the ODE system as follows:

$$\begin{aligned} \frac{du_1}{dt} &= f_1(t, u_1, u_2, u_3), \\ \frac{du_2}{dt} &= f_2(t, u_1, u_2, u_3), \\ \frac{du_3}{dt} &= a u_1 u_2 + b, \end{aligned} \tag{14}$$

where a, b are unknown constants to be inferred and f_1, f_2 are unknown functions, whose analytic regressions we seek to discover. Here, we choose to use neural networks (NNs) with two hidden layers, each of which equipped with 50 neurons, and hyperbolic tangent activation. Following the proposed method, we use one NN as the surrogate of u, which has 1-dimensional input and 3-dimensional output, and one NN to approximate concatenation of f_1 and f_2, which has 4-dimensional input and 2-dimensional output. The results are shown in Tables 3 and 4. As we can see, both steps of our approach yield accurate inferences of a, b and f_1, f_2.

	a	b
Inference	-1.9975	0.0002
Exact	-2	0

Table 3: Inference of unknown constants using PINNs.

We reiterate that a and b are jointly inferred in the PINNs step while the identification of f_1 and f_2 is done separately in the symbolic regression step. We also note that results from symbolic regression may be sensitive to the metrics of model selection, to the candidates of unary and binary operators, and to the complexities of the operators. In this example, the metric is the score defined in [8] and described in Section 2.2, the binary operators are chosen to be addition and multiplication, the unary operators are identity, sin, cos, exponential and reciprocal, and the complexities of unary operators are set to 1, 3, 3, 3, 3, respectively.

f_1	$u_2u_3e^{-0.0986t}$	$0.6694u_2u_3$	u_2u_3
Score	3.8855	0.6695	0.5910
f_2	u_1u_3	$u_1u_3 + \frac{0.0014}{t+0.0314}$	$u_1u_3 - 0.0020 + \frac{0.0017}{t+0.0379}$
Scores	2.9805	0.5552	0.0617

Table 4: Identifications of the dynamics of u_1 and u_2 using PINNs and symbolic regression. Top 3 identifications, in terms of the score defined in Section 2.2, are shown.

B Effect of sample size on inference quality

In the simulated dataset, we are provided with training data $c(t)$ at $t = 0, 1, \cdots, T$. In the primary exposition, we have determined that a value of $T = 2$ suffices for the purpose of training, inference, and projection. This conclusion is, in part, attributable to the additional physical constraint that we have imposed, namely, $f(0) = f(1) = 0$. In the absence of these constraints, our investigation has revealed that a greater amount of data is required to optimize the model effectively. To illustrate this finding, we have conducted three simulations with $T = 2, 4, 6$ and have plotted the inference results obtained using PINNs and the symbolic model, as depicted in Figure 7. It is noteworthy that in all cases, the parameters κ and α are inferred correctly; yet, the parametric function f_ϕ does not match f in areas where no data for c is available. This is particularly pronounced for larger concentration c values, which correspond to subjects with tau concentration levels close to 0 in the initial stages. The rightmost column of Figure 7 shows that the inferred $f(c)$ does not agree well with the ground truth. Interestingly, when $T = 4$ and $T = 6$, the symbolic regression model f_{sym} discovered from the neural network f_ϕ outperforms f_ϕ in approximating the underlying target. One possible explanation for this phenomenon is that f_{sym} is trained with the concentration c from the training set of f_ϕ, which means that the region that corresponds to the incorrect inference by f_ϕ is not fed into the symbolic regression model. As a result, the underlying inductive bias directs f_{sym} towards the correct solution automatically. Finally, we note that Table 5 demonstrates that, to obtain the correct symbolic model when there are no constraints on the boundary conditions of f, data up to $T = 6$ is required.

	Group 1	Group 2	Group 3	Group 4
$T = 2$	$2.88ce^{-1.05e^c}$	$c(-1 + 2e^{-0.55c^2})$	$c^2(1.76 - 0.78e^c)$	$0.37ce^{-0.71c^2}$
$T = 4$	$c(-1.16c + 0.12e^c + 0.9)$	$c(1 - c^2)$	$c^2(1 - c)$	$c(0.51c - 0.49e^c + 0.86)$
$T = 6$	$c(1 - c)$	$c(1 - c^2)$	$c^2(1 - c)$	$c(0.37 - 0.37c)e^c$

Table 5: **Inferred reaction terms when $T = 2, 4, 6$, without fixed boundary conditions for f.** When boundary conditions for f are not fixed, more data are required to guarantee the correct inference result by symbolic regression model. The more data we are provided, the better inference quality will be achieved. To get f_{sym} correct in 4 groups, we need data up to $T = 6$.

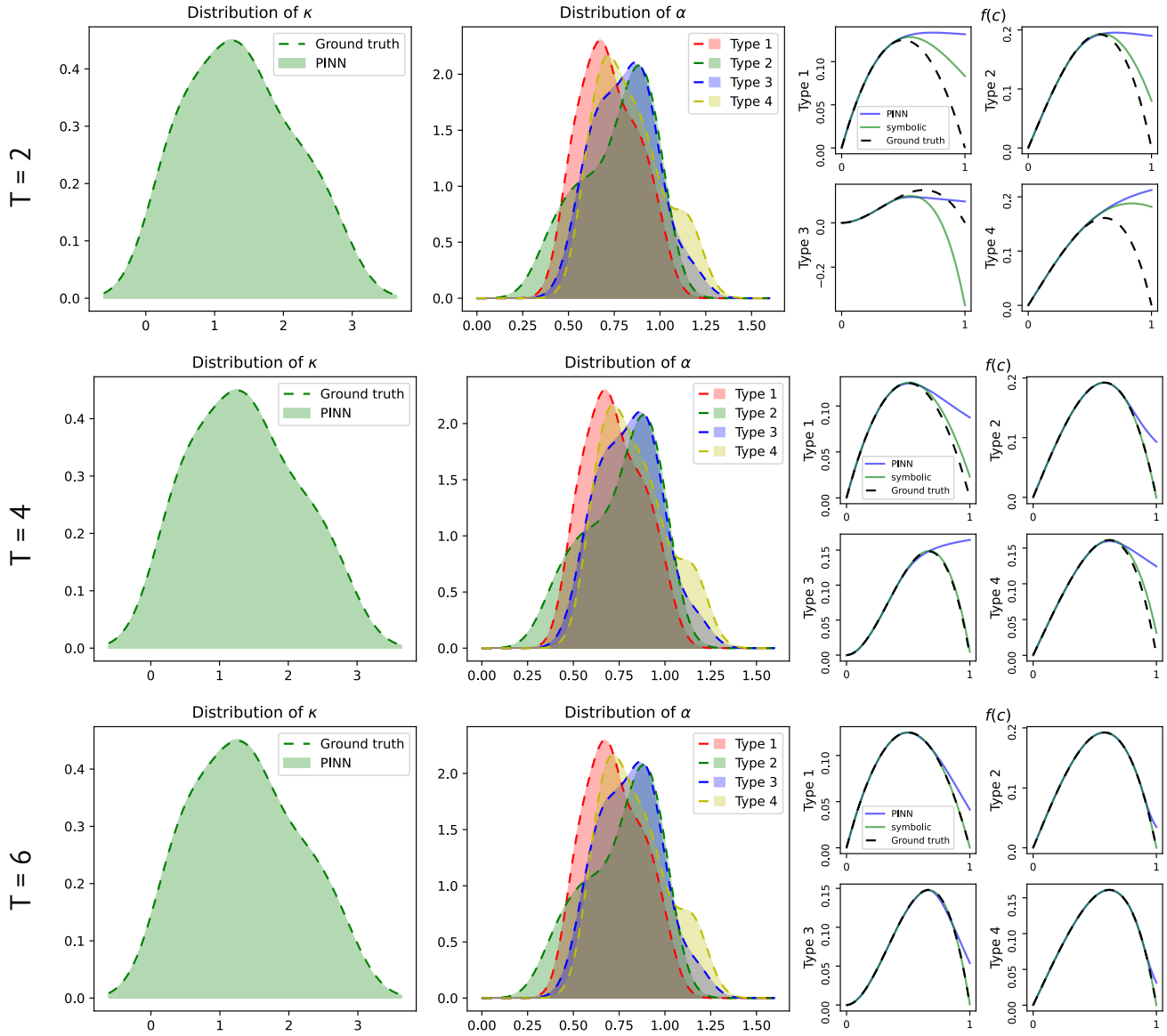

Figure 7: **PINN inference results when $T = 2, 4, 6$, without fixed boundary conditions for f.** While the learned parameters κ and α perform well in all instances, the parametric function f_ϕ fails to align with f in regions where there is no available data for c.

Acknowledgements

We would like to thank Dr. Amelie Schäfer and Dr. Kevin Linka for helpful discussions. ZZ, ZZ and GEK were supported by the DOE SEA-CROGS project (DE-SC0023191), the MURI-AFOSR FA9550-20-1-0358 project, and the ONR Vannevar Bush Faculty Fellowship (N00014-22-1-2795). EK was supported by the NSF CMMI grant 2320933 Automated Model Discovery for Soft Matter.

References

[1] Keith A Johnson, Aaron Schultz, Rebecca A Betensky, J Alex Becker, Jorge Sepulcre, Dorene Rentz, Elizabeth Mormino, Jasmeer Chhatwal, Rebecca Amariglio, Kate Papp, et al. Tau positron emission tomographic imaging in aging and early a lzheimer disease. *Annals of neurology*, 79(1):110–119, 2016.

[2] Julio C Rojas and Adam L Boxer. Targeting tauopathies for therapeutic translation. *Nature Reviews Neurology*, 12(2):74–76, 2016.

[3] Mattia Corti, Francesca Bonizzoni, Paola F Antonietti, and Alfio M Quarteroni. Uncertainty quantification for fisher-kolmogorov equation on graphs with application to patient-specific alzheimer disease. *arXiv preprint arXiv:2305.03619*, 2023.

[4] Amelie Schäfer, Mathias Peirlinck, Kevin Linka, Ellen Kuhl, and Alzheimer's Disease Neuroimaging Initiative (ADNI). Bayesian physics-based modeling of tau propagation in Alzheimer's disease. *Frontiers in Physiology*, 12:702975, 2021.

[5] Amelie Schäfer, Pavanjit Chaggar, Alain Goriely, Ellen Kuhl, and Alzheimer's Disease Neuroimaging Initiative. Correlating tau pathology to brain atrophy using a physics-based bayesian model. *Engineering with Computers*, 38(5):3867–3877, 2022.

[6] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[7] Miles Cranmer. Interpretable machine learning for science with PySR and SymbolicRegression. jl. *arXiv preprint arXiv:2305.01582*, 2023.

[8] Miles Cranmer, Alvaro Sanchez Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *Advances in Neural Information Processing Systems*, 33:17429–17442, 2020.

[9] Bastian Zapf, Johannes Haubner, Miroslav Kuchta, Geir Ringstad, Per Kristian Eide, and Kent-Andre Mardal. Investigating molecular transport in the human brain from mri with physics-informed neural networks. *Scientific Reports*, 12(1):15475, 2022.

[10] Zongren Zou, Xuhui Meng, Apostolos F Psaros, and George Em Karniadakis. NeuralUQ: A comprehensive library for uncertainty quantification in neural differential equations and operators. *arXiv preprint arXiv:2208.11866*, 2022.

[11] Ehsan Kharazmi, Min Cai, Xiaoning Zheng, Zhen Zhang, Guang Lin, and George Em Karniadakis. Identifiability and predictability of integer-and fractional-order epidemiological models using physics-informed neural networks. *Nature Computational Science*, 1(11):744–753, 2021.

[12] Minglang Yin, Zongren Zou, Enrui Zhang, Cristina Cavinato, Jay D Humphrey, and George Em Karniadakis. A generative modeling framework for inferring families of biomechanical constitutive laws in data-sparse regimes. *arXiv preprint arXiv:2305.03184*, 2023.

[13] Kevin Linka and Ellen Kuhl. A new family of Constitutive Artificial Neural Networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403:115731, 2023.

[14] Kevin Linka, Sarah R St Pierre, and Ellen Kuhl. Automated model discovery for human brain using Constitutive Artificial Neural Networks. *Acta Biomaterialia*, 160:134–151, 2023.

[15] Kevin Linka, Adrian Buganza Tepole, Gerhard A Holzapfel, and Ellen Kuhl. Automated model discovery for skin: Discovering the best model, data, and experiment. *Computer Methods in Applied Mechanics and Engineering*, 410:116007, 2023.

[16] Andreï N Kolmogorov. A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. *Moscow University Bulletin of Mathematics*, 1:1–25, 1937.

[17] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369, 1937.

[18] Alan C Newell and John A Whitehead. Finite bandwidth, finite amplitude convection. *Journal of Fluid Mechanics*, 38(2):279–303, 1969.

[19] Lee A Segel. Distant side-walls cause slow amplitude modulation of cellular convection. *Journal of Fluid Mechanics*, 38(1):203–224, 1969.

[20] Yakov B Zeldovich and David A Frank-Kamenetskii. The theory of thermal propagation of flames. *Zh. Fiz. Khim*, 12:100–105, 1938.

[21] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[22] Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. DeepXDE: A deep learning library for solving differential equations. *SIAM review*, 63(1):208–228, 2021.

[23] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-PINNs: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

[24] Ameya D Jagtap, Ehsan Kharazmi, and George Em Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020.

[25] Guofei Pang, Lu Lu, and George Em Karniadakis. fPINNs: Fractional physics-informed neural networks. *SIAM Journal on Scientific Computing*, 41(4):A2603–A2626, 2019.

[26] Zhiping Mao, Ameya D Jagtap, and George Em Karniadakis. Physics-informed neural networks for high-speed flows. *Computer Methods in Applied Mechanics and Engineering*, 360:112789, 2020.

[27] Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (PINNs) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738, 2021.

[28] Zongren Zou and George Em Karniadakis. L-HYDRA: Multi-Head Physics-Informed Neural Networks. *arXiv preprint arXiv:2301.02152*, 2023.

[29] Zongren Zou, Xuhui Meng, and George Em Karniadakis. Correcting model misspecification in physics-informed neural networks (PINNs). *arXiv preprint arXiv:2310.10776*, 2023.

[30] Lena Podina, Brydon Eastman, and Mohammad Kohandel. A PINN approach to symbolic differential operator discovery with sparse data. *arXiv preprint arXiv:2212.04630*, 2022.

[31] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

[32] Amelie Schäfer, Elizabeth C Mormino, and Ellen Kuhl. Network diffusion modeling explains longitudinal tau pet data. *Frontiers in Neuroscience*, 14:566876, 2020.

[33] Jennifer A McNab, Brian L Edlow, Thomas Witzel, Susie Y Huang, Himanshu Bhat, Keith Heberlein, Thorsten Feiweier, Kecheng Liu, Boris Keil, Julien Cohen-Adad, et al. The human connectome project and beyond: initial applications of 300 mt/m gradients. *Neuroimage*, 80:234–245, 2013.

[34] M X Henderson, E J Cornblath, A Darwich, B Zhang, H Brown, R J Gathagan, R M Sandler, D S Bassett, T J Trojanowski, and Lee V M Y. Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis. *Nature Neuroscience*, 22(8):1248–1257, 2019.

[35] Ellen Kuhl. Connectomics of neurodegeneration. *Nature Neuroscience*, 22(8):1199–1202, 2019.

[36] Sveva Fornari, Amelie Schäfer, Mathias Jucker, Alain Goriely, and Ellen Kuhl. Prion-like spreading of alzheimer's disease within the brain's connectome. *Journal of the Royal Society Interface*, 16(159):20190356, 2019.

[37] Xiaoliang Wan and George Em Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal on Scientific Computing*, 28(3):901–928, 2006.

[38] Paula Chen, Tingwei Meng, Zongren Zou, Jérôme Darbon, and George Em Karniadakis. Leveraging Multi-time Hamilton-Jacobi PDEs for Certain Scientific Machine Learning Problems. *arXiv preprint arXiv:2303.12928*, 2023.

[39] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[40] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.

[41] Ying H Huang, Zheng Xu, Cheng Qian, and Li Liu. Solving free-surface problems for non-shallow water using boundary and initial conditions-free physics-informed neural network (bif-PINN). *Journal of Computational Physics*, 479:112003, 2023.

[42] S Berrone, C Canuto, M Pintore, and N Sukumar. Enforcing Dirichlet boundary conditions in physics-informed neural networks and variational physics-informed neural networks. *arXiv preprint arXiv:2210.14795*, 2022.

[43] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, 30, 2017.

[44] Apostolos F Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, page 111902, 2023.

[45] Balázs Szalkai, Csaba Kerepesi, Bálint Varga, and Vince Grolmusz. Parameterizable consensus connectomes from the human connectome project: the budapest reference connectome server v3. 0. *Cognitive neurodynamics*, 11:113–116, 2017.

[46] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[47] Simon S Du, Xiyu Zhai, Barnabas Poczos, and Aarti Singh. Gradient descent provably optimizes over-parameterized neural networks. *arXiv preprint arXiv:1810.02054*, 2018.

[48] Chaoyue Liu, Libin Zhu, and Mikhail Belkin. Loss landscapes and optimization in over-parameterized non-linear systems and neural networks. *Applied and Computational Harmonic Analysis*, 59:85–116, 2022.