



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Probabilistic Temporal Prediction of Continuous Disease Trajectories and Treatment Effects Using Neural SDEs

Joshua Durso-Finley^{*,1,4}, Berardino Barile^{*,1,4}, Jean-Pierre Falet^{3,4}, Douglas L. Arnold^{1,3}, Nick Pawlowski², Tal Arbel^{1,4}

¹Center for Intelligent Machines, McGill University. ²Microsoft Research. ³Montreal Neurological Institute, McGill University. ⁴MILA (Quebec AI institute).

* These authors contributed equally

Abstract. Personalized medicine based on medical images, including predicting future individualized clinical disease progression and treatment response, would have an enormous impact on healthcare and drug development, particularly for diseases (e.g. multiple sclerosis (MS)) with long-term, complex, heterogeneous evolutions and no cure. In this work, we present the first stochastic causal temporal framework to model the continuous temporal evolution of disease progression via Neural Stochastic Differential Equations (NSDE). The proposed causal inference model takes as input the patient’s high dimensional images (MRI) and tabular data and predicts both factual and counterfactual progression trajectories on different treatments in latent space. The NSDE permits the estimation of high-confidence personalized trajectories and treatment effects. Extensive experiments were performed on a large, multi-centre, proprietary dataset of patient 3D MRI and clinical data acquired during several randomized clinical trials for MS treatments. Our results present the first successful uncertainty-based causal Deep Learning (DL) model to: (a) accurately predict future patient MS disability evolution (e.g. EDSS) and treatment effects leveraging baseline MRI and (b) permit the discovery of subgroups of patients for which the model has high confidence in their response to treatment even in clinical trials which did not reach their clinical endpoints.

1 Introduction

Consider patients with long-term, incurable disease, characterised by complex and heterogeneous disease courses and treatment responsiveness. In multiple sclerosis (MS), a common disabling neurological condition, no strongly predictive imaging marker exists for progressive disability evolution [23], and it’s unclear if available drugs could be effective at slowing disability progression in subsets of patients, as their efficacy at the population level has been limited. In these contexts, the impact of a deep learning (DL) clinical decision support tool that can predict the future continuous disease evolution for individual patients based

on (non-invasive) medical scans, and standard clinical and demographic information, would be enormous. Even more impactful would be if it could predict the relative effect of available treatments compared to an untreated course for *this patient*, and quantify the degree of confidence in its predictions for safe deployment in high-risk settings. Such is the promise of DL for image-based precision medicine. Unfortunately, this promise has not yet been met [26].

Predictions of clinical disease outcomes have, for the most part, focused on outcomes at fixed future timepoints[38, 14, 21]. In MS, some have proposed to model particular radiomic feature evolutions and link them to clinical outcomes ([29, 13]). Learning the dynamics of long-term disease evolution for *clinical* endpoints based on medical images would be hugely beneficial to patient care in general. This would lead to a better understanding of disease evolution across the population, including the discovery of new personalized predictive image markers, thereby providing the basis for more equitable patient care. Furthermore, forecasting continuous future clinical outcome trajectories based on early observations, especially with irregularly spaced, missing and noisy observations, would be enormously impactful for improved patient care and clinical trial analysis.

Recently, Neural Ordinary Differential Equations (NODEs) [7] were proposed to improve upon discrete-time models based on RNNs [35]. NODEs were applied to forecasting in clinical settings but only with tabular or synthetic input data [22, 37, 10, 27]. Given that measured clinical outcomes can be noisy (e.g. Expanded Disability Status Scale [32] in MS), and that DL models are not always accurate, probabilistic temporal models that represent a distribution over possible trajectories would be desirable and help quantify the uncertainties in the model’s predictions. Although many approaches for estimating uncertainty have been proposed for DL predictions in medical imaging [1, 2], applications in personalized medicine have been restricted to single time-point predictions [12]. Neural stochastic differential equations (NSDEs)[42] have recently been proposed as an extension of NODE for probabilistic modeling of trajectories in latent space for tabular data [30]. However, to our knowledge, NSDE-based methods have not been applied to medical image analysis. Finally, in chronic conditions where many treatments are available, predicting a future individual treatment effect (ITE) requires estimating the relative effect of one treatment against a comparator. Causal inference for prediction of future ITE in medical image analysis remains in its infancy [33], with some recent work focused on future fixed-timepoint outcomes[11, 41]. The importance of causality for image-based personalised medicine warrants extending prior work on causal modeling for ITE estimation to NSDEs, which have not been studied in this context.

In this paper, we present the first probabilistic temporal causal model for image-based personalized medicine. The encoder combines the pre-treatment (baseline) high-dimensional images (i.e. MRI) with tabular clinical data, and projects its latent representation forward in time to model the continuous evolution of clinical trajectories via an NSDE. The model infers ITE by comparing the predicted factual and counterfactual future clinical disease trajectories for

patients on all available treatments in latent space and additionally uses the stochastic nature of the NSDE for sampling-based variance estimation to quantify uncertainty in the predicted trajectory (and the ITE on the trajectory). Extensive experiments are performed on a large (over 3600 patients), multi-centre, proprietary dataset of 3D MRI acquired from patients with MS during six randomized clinical trials (RCTs), testing seven active treatments, and two placebo groups. We illustrate results on two different subtypes of MS: relapsing remitting MS (RRMS) and secondary progressive (SPMS). Compared to a discrete temporal model (LSTM) and a fixed time-point Encoder-Decoder regressor (see Supplemental Materials, Figure 2 for details), our model more accurately predicts future EDSS. Moreover, we find that sampling-based uncertainty estimates allow for isolating subsets of patients for which the model is more confident, improving the predictive accuracy of the trajectory and ITE.

2 Method

A causal stochastic temporal model is proposed: At baseline, multi-dimensional MRI, and the associated tabular data (clinical, demographic, and subtype), are encoded in two distinct lower-dimensional latent spaces. The treatment label, one-hot-encoded, is concatenated together with the tabular and image embeddings and fed to the NSDE which projects the combined latent representation (from baseline) forward in time. Finally, a decoder is trained to map the predicted latent encoding to the outcome of interest (e.g. future EDSS) (see Figure 1). The general mathematical framework is now described.

2.1 Learning Probabilistic Trajectories via Stochastic Neural ODE

Let $t \in \mathbb{R}$ be time where t_0 represents the time of the initial visit. Let $X(t_0) = \{x_i(t_0) \in \mathbb{R}^{(c_1, c_2, c_3)}\}_{i=1}^n$ be the 3D MRI at time $t = 0$, i the patient index, (c_1, c_2, c_3) the image size and $W(t_0) = \{w_i(t_0) \in \mathbb{R}^d\}_{i=1}^n$ be the associated clinical data. $X(t_0)$ and $W(t_0)$ are mapped to a lower dimensional representation $Z_x(t_0) = \phi(X(t_0); \theta_\phi) \in \mathbb{R}^{h_1}$ and $Z_w(t_0) = \psi(W(t_0); \theta_\psi) \in \mathbb{R}^{h_2}$, where ϕ and ψ are trainable functions parametrized by θ . Let $Z(t_0)$ be the combined latent representation obtained by concatenating the two latent vectors such that $Z(t_0) \in \mathbb{R}^{h_1+h_2}$. For each time series, we want to predict the clinical target $Y_i(t | S = s)$ at some timepoint $t > t_0$ in the future, condition to treatment $S = s$. We use the NSDE to model a stochastic process for Z whose dynamics are derived from a deterministic (drift) function f_θ and a stochastic (diffusion) component g_θ :

$$\frac{dZ(t)}{dt} = f_\theta(Z(t) | S = s) + g_\theta(Z(t) | S = s) \circ d\Omega_t \quad (1)$$

where $d\Omega_t \sim \mathcal{N}(0, \Delta t I_w)$ represents the h_1+h_2 -dimensional Brownian motion. In our model, f_θ and g_θ are parameterized by neural networks. The latent encoding

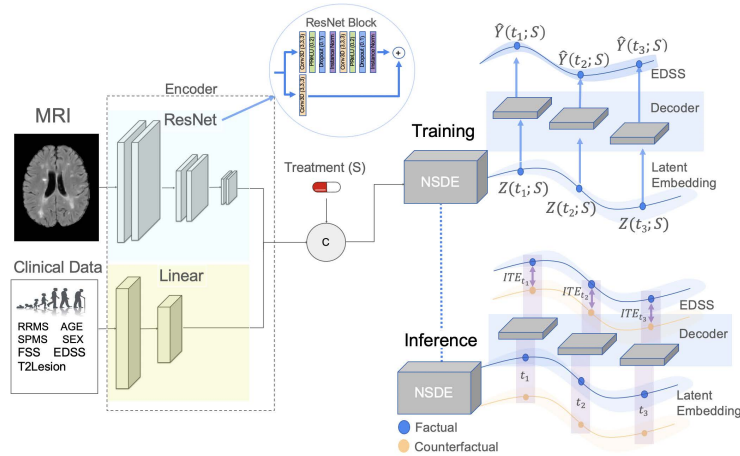


Fig. 1: Overall framework: 3D MRI and tabular data (clinical, demographic, and subtypes) are encoded, concatenated and passed to the NSDE solver. During supervised **Training** (end-to-end), the model learns to represent the temporal trajectory in latent space conditioned on the treatment the patient received. The decoder maps the predicted longitudinal latent representations into outcomes (EDSS scores). During **Inference**, the model projects the embeddings onto the latent space forwarded in time and conditioned on the treatment (or control), permitting factual and counterfactual latent trajectories and uncertainties to be estimated. The decoder predicts future outcomes, and uncertainties, at any point along the continuous latent trajectories. ITE, and associated uncertainties, can be estimated by comparing probabilistic trajectories on and off treatments.

for a future latent state, given treatment s and initial state $Z(t_0)$, can be obtained through Stratonovich integration [20] as follows:

$$Z_{i,j}(t) = \text{ODESolve}(Z_i(t_0), f_\theta(\cdot) + g_\theta(\cdot) \circ \frac{d\Omega_{t,j}}{dt} \mid S = s, [t_0, t]). \quad (2)$$

Thus, $Z_{i,j}(t)$ represents the forecasted latent space for patient “ i ”, with Brownian motion sample “ j ”, at future time “ t ”, and conditioned on the treatment assignment $S = s$. Lastly, the latent state is decoded through a deterministic function parameterized by a neural network $\Phi(Z_{ij}(t) \mid S = s; \theta_\Phi) \in \mathbb{R}$ to obtain predicted outcome $\hat{Y}_{ij}(t \mid S = s)$.

2.2 Temporal Counterfactual Trajectories and ITE

In our proposed framework, we extend individual treatment effect (ITE) [34] estimation to the longitudinal setting, by comparing latent *trajectories* of patients on and off treatments from baseline inputs. Following the Neyman/Rubin Potential Outcome Framework [18], we consider two potential outcomes y_0 (control/placebo) and y_s (treatment s). For patient i with triples $\mathcal{D} = \{(z_{i,t}, y_{i,t}, s_{i,t})\}_{i=1}^n$

where $t > t_0$, the ITE can be defined as a function of time: $\text{ITE}_s(t) = y_s(t) - y_0(t)$, where $y_s(t)$ and $y_0(t)$ represent *potential* outcomes for treatment and control patients, respectively. The $\text{ITE}_s(t)$ is an unobservable causal estimand, as only one of the two potential outcomes can be observed. For each patient i and Brownian sample j , the following causal estimand for ITE, conditioned on the latent encoding Z at timepoint t , is used: $\tau_s^{(ij)}(z; t) = \mathbb{E}[y_s | z; t]_{ij} - \mathbb{E}[y_0 | z; t]_{ij}$. In the case of RCTs (our case), $(Y_0, Y_s) \perp S | Z(t_0)$, and therefore we can learn an unbiased estimator for τ_s [28]. A trajectory ITE can thus be estimated as follows:

$$\bar{\tau}_s^{(ij)} = \frac{1}{t_k - t_0} \sum_{\forall t \geq t_0} \tau_s^{(ij)}(z; t) \quad (3)$$

where k is the number of longitudinal timepoints. As the NSDE permits stochastic sampling of predicted factual and counterfactual trajectories, we can integrate over j and obtain the mean, $\mu_s^{(i)}(z)$, and variance, $\sigma_s^{(i)}(z)$, of the longitudinal ITE for patient i (Eq. (3)). The variance can thus be used as an uncertainty estimand for the longitudinal ITE.

3 Experiments and Results

3.1 Dataset

The data is pooled from six randomized clinical trials (RCTs). Five trials, BRAVO [40], OPERA 1 [16], OPERA 2 [16], DEFINE [4], and ADVANCE [6] enrolled patients with relapsing-remitting MS (RRMS) and randomized them to one of the following treatments: **Placebo** ($n = 455$), **Laquinimod** ($n = 295$), **INterFeron Beta-1a IntraMuscular (INFB-IM)** ($n = 326$), **INterFeron Beta-1a SubCutaneous (INFB-SC)** ($n = 572$), **DiMethyl Fumarate (DMF)** ($n = 225$), **Ocrelizumab** ($n = 554$), **Pegylated-Interferon (PINFB)** ($n=840$). ASCEND [19] randomized secondary progressive MS (SPMS) patients to **Placebo** ($n = 302$) and **Natalizumab** ($n = 293$) (See Supplemental Material, Table 1 for more dataset statics.) The primary clinical target of interest is the EDSS score, which measures rate of disability. It is an ordinal rating scale ranging from 0 (no disability) to 10 (death) in steps of 0.5. Inputs to the model consisted of: (1) clinical and demographic features available at baseline and pre-trial screening visits: age, sex, functional scores (FSS), T2 lesion volume and EDSS. (2) Fluid Attenuated Inverse Recovery (FLAIR) MRI acquired at baseline at $1 \times 1 \times 3$ mm resolution. The MRI data underwent several preprocessing steps to enforce homogeneity among the trials [9, 8, 5, 24, 36] (See Supplemental Material for a full description). Follow-up EDSS scores, measured inconsistently over time at intervals of approximately 12 weeks, up until 96 weeks, were used as regression targets.

3.2 Experimental and Implementation Details

A ResNet encoder [17] with 3 residual blocks, and a feed-forward network with 2 linear layers and ELU activations, were used to encode the 3D MRI and the

associated clinical data, respectively. Balanced MSE [31] was used as the regression loss. The entire model was trained end-to-end and evaluated as part of a nested 4×4 -fold Cross-Validation (CV) strategy. Specifically, the dataset was initially split into 4 folds. At each iteration, each fold was used as a held-out test set, while the remaining 3 folds were used for training, validation and hyperparameter tuning. The process was repeated until each of the 4-folds was used as hold-out test set. Results are presented based on CV aggregation or crogging [3], to improve the generalization error estimate. During training the model learns from observed factual outcomes. During inference, estimates for the counterfactual trajectories and ITE are obtained by sampling trajectories conditioned on specific treatments.

3.3 Evaluating Predicted EDSS Trajectories and Uncertainties

Evaluations on the factual predictions for future EDSS scores are performed. Fig.2(a) reports the MSE between available and predicted EDSS scores. These are provided at (approximately) 12-week intervals up until week 96. To estimate uncertainty, the NSDE was run 30 times with different Brownian noise samples for each prediction. For all patients, the MSE was below 1 at all timepoints, demonstrating that the model accurately predicts disease evolution up to 2 years in the future (e.g. [27] for comparison). As expected, more accurate results were obtained earlier in the trajectory, with reduced performance farther into the future.

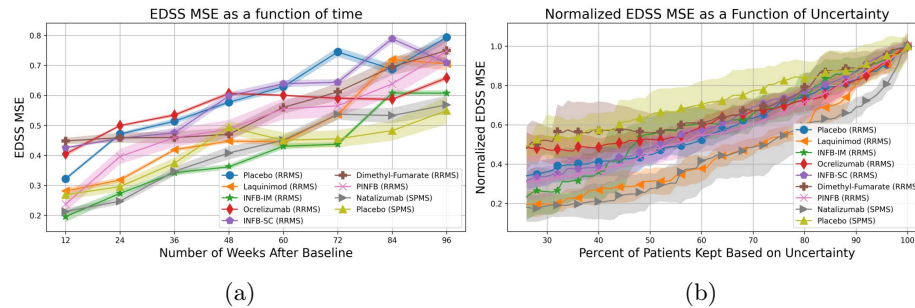


Fig. 2: (a) MSE of factual EDSS predictions as a function of time. Evaluations are shown at 12-week intervals up to week 96. Results are shown for 7 active treatments and 2 placebo control groups (RRMS and SPMS). (b) Normalized Mean Squared Error (MSE) of factual EDSS predictions for patients kept based on model uncertainties. One can see a clear decrease in error when moving from right (100% of the patients) to left (30% of the patients with the most confident predictions). Shaded areas show the variance from different sets of outer fold aggregation.

The proposed approach was compared with different baselines: (1) A Long-Short-Term-Memory (LSTM) model was trained, where irregular visits were grouped together to the closest 12-week interval and (2) A fixed time-point Encoder-Decoder regressor. (The architectural details of the models are found in the Supplemental Materials.) Results are found in Table 1. The proposed NSDE framework had an overall MSE of 0.575 (0.012), as compared to 0.674 (0.014) (non-temporal) and 0.662 (0.014) (LSTM) averaged over all treatments.

Treatment	Encoder-Decoder	LSTM	NSDE
Laquinimod	0.769 (0.022)	0.691 (0.006)	0.672 (0.008)
INFB-SC	0.624 (0.004)	0.614 (0.010)	0.590 (0.006)
Ocrelizumab	0.603 (0.008)	0.616 (0.012)	0.555 (0.003)
INFB-IM	0.737 (0.009)	0.712 (0.010)	0.699 (0.010)
Dimethyl Fumarate	0.708 (0.015)	0.690 (0.005)	0.679 (0.011)
PINFB	0.572 (0.005)	0.584 (0.004)	0.522 (0.008)
Natalizumab	0.882 (0.024)	0.857 (0.015)	0.533 (0.013)
Placebo (SPMS)	0.752 (0.013)	0.675 (0.021)	0.358 (0.013)
Placebo (RRMS)	0.672 (0.006)	0.696 (0.008)	0.655 (0.006)

Table 1: Comparison of Factual Results across Baselines: Overall Mean and standard error, computed as the average MSE for the EDSS outcomes across timepoints, and across patients, for three models: (1) A fixed time-point Encoder-Decoder regressor, (2) LSTM, and (3) proposed NSDE. The proposed NSDE model generally outperforms the other models in all drugs and placebo groups. See Supplemental Materials for model details.

Evaluation of the uncertainty estimates is performed in order to confirm that higher confidence estimates correlate with lower errors. To this end, patients were ordered based on the model confidence (see Section 2.2). Fig.2 reports the MSE results averaged along the entire trajectories for patients filtered based on their uncertainty estimates. To better analyze the impact of uncertainty on error, the MSE was normalized to the baseline case where no uncertainty filtering was performed (100% of the data). Results indicate that trajectory predictions for all treatments decrease in error when more confident predictions are considered, confirming the desired performance of the predicted uncertainties. For example, when considering the 30% most confident predictions (x-axis), the normalized MSE score was at 0.18 for Natalizumab, which was 83% lower than the reference case (no-uncertainty).

3.4 Counterfactual Predictions and Treatment Response

The proposed model predicts future individual counterfactual (CF) trajectories of patient disability scores (EDSS), resulting from a causal intervention on the treatment variable. Given that validating CFs is challenging in the absence of

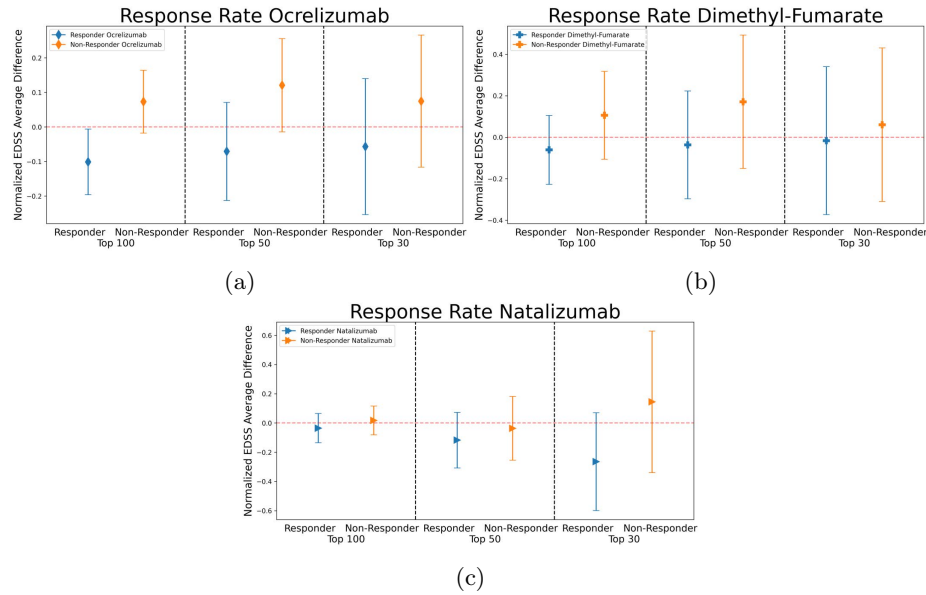


Fig. 3: Treatment response estimates (uplift) for three drugs (Ocrelizumab, DMF, Natalizumab). Comparison of predicted **responders** and **non-responders** at three different levels of uncertainties (30%, 50% and 100%-no uncertainty). Negative values indicate response through a reduction in average EDSS change on the drug as compared to placebo. Our model finds subgroups of responders for these drugs, known to be of moderate and high efficacy. The other drugs are considered to be low efficacy (Results in Supplemental Materials).

ground truth, evaluations are based on *uplift* scores [39], defined as the average difference in predicted EDSS change between treated and counterfactual (non-treated) patients. When sorting patients by their uplift score, predicted **responders**, and **non-responders** can be identified. Responders are expected to have negative uplift values, indicating that the change in EDSS is higher on placebo than it is on treatment. Figure 3 shows three examples of uplift plots for SPMS patients treated with Natalizumab, and RRMS patients treated with either DMF or Ocrelizumab. These were chosen because they are known to be moderate and high efficacy drugs[25, 15]. Results are depicted for three levels of uncertainties in order to explore whether the identification of responders to treatment can be further improved by considering the subset of patients with high predicted confidence. The results indicate that uplift scores are generally negative (lower) for the responder group (33% of patients with highest uplift) compared to the non-responder (33% of patients with lowest uplift), suggesting that the model is able to identify responders with high degree of accuracy. When comparing the 30% of patients with the highest confidence predictions, the

gap between responders and non-responder widens, suggesting an even greater treatment effect for these patients.

4 Conclusions

For chronic neurological diseases such as MS, the ability to accurately predict continuous individual trajectories of disease evolution on different treatments from baseline MRI would be of substantial importance to improving patient care and drug development. In this work, a probabilistic temporal model is proposed for (1) predicting continuous disability trajectories in latent space based on NSDEs, and (2) future individual treatment responses, and validated on a large (>3600 patients) unified RCT dataset for a variety of treatments. Causal inference, combined with uncertainty estimates, provides additional trustworthiness in predicted counterfactual trajectories, and permit the identification of high confidence subgroups of highly responsive patients. Future work will explore other strategies to measure treatment response and uncertainty estimation over trajectories. Finally, clustering of subgroups of responders will permit the discovery of common predictive imaging markers across the population.

Acknowledgements: This investigation was supported by the International Progressive Multiple Sclerosis Alliance (PA-1412-02420) and the companies who generously provided the data: Biogen, BioMS, MedDay, Novartis, Roche / Genentech, and Teva, the Canada Institute for Advanced Research (CIFAR) AI Chairs program, the Natural Sciences and Engineering Research Council of Canada, Microsoft Research, Mila - The Quebec AI Institute, the MS Society of Canada, and Calcul Quebec, and the Digital Research Alliance of Canada. The authors would like to thank L. Collins and M. Dadar for preprocessing the MRI data, Z. Caramanos, A. M. Pinzon, C. Guttman and I. Mórocz for collating the clinical data, and S. Narayanan and M.-P. Sormani for their MS expertise.

Disclosure of Interests: The authors have no competing interests to declare.

References

1. Abdar, M., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion* (2021)
2. Ashby, D.: Bayesian statistics in medicine: A 25 year review. *Statistics in medicine* (2006)
3. Barrow, D.K., Crone, S.F.: Crogging (cross-validation aggregation) for forecasting — A novel algorithm of neural network ensembles on time series subsamples. In: *The 2013 International Joint Conference on Neural Networks (IJCNN)* (2013)
4. Bomprezzi, R.: Dimethyl fumarate in the treatment of relapsing–remitting multiple sclerosis: an overview. *Therapeutic Advances in Neurological Disorders* (2015)
5. Boyes, R.G., et al.: Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage* (2008)
6. Calabresi, P.A., et al.: Pegylated interferon beta-1a for relapsing-remitting multiple sclerosis (ADVANCE): a randomised, phase 3, double-blind study. *The Lancet Neurology* (2014)
7. Chen, T., et al.: Neural Ordinary Differential Equations. *CoRR* (2018)
8. Collins, D.L., et al.: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* (1994)
9. Collins, L., et al.: Animal: Validation and Applications of Nonlinear Registration-Based Segmentation. *IJPRAI* (2011)
10. De Brouwer, E., et al.: GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series. *Advances in Neural Information Processing Systems* (2019)
11. Durso-Finley, J., et al.: Personalized Prediction of Future Lesion Activity and Treatment Effect in Multiple Sclerosis from Baseline MRI. *MICCAI* (2022)
12. Durso-Finley, J., et al.: Improving Image-Based Precision Medicine with Uncertainty-Aware Causal Models. *MICCAI* (2023)
13. Eshaghi, A., et al.: Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nature Communications* (2021)
14. Fang, C., et al.: Deep learning for predicting COVID-19 malignant progression. *Medical Image Analysis* (2021)
15. Gold, R., et al.: Safety and efficacy of delayed-release dimethyl fumarate in patients with relapsing-remitting multiple sclerosis: 9 years’ follow-up of DEFINE, CONFIRM, and ENDORSE. *her Adv Neurol Disord* (2020)
16. Hauser, S.L., et al.: Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *New England Journal of Medicine* (2017)
17. He, K., et al.: Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)
18. Imbens, G., Rubin, D.: *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press (2015)
19. Kapoor, R., et al.: Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): A phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension. *The Lancet Neurology* (2018)
20. Kidger, P., et al.: Neural SDEs as Infinite-Dimensional GANs. *International Conference on Machine Learning* (2021)
21. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Medical image analysis* (2017)
22. Lu, J., et al.: Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience* (2021)

23. Lucchinetti, C., et al.: Heterogeneity of multiple sclerosis lesions: Implications for the pathogenesis of demyelination. *Annals of neurology* (2000)
24. Manjón, J.V., et al.: Adaptive non-local means denoising of MR images with spatially varying noise levels. *J Magn Reson Imaging* (2010)
25. Montalban, X., et al.: Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med.* (2017)
26. Naji, Y., et al.: Artificial Intelligence and Multiple Sclerosis: Up-to-Date Review. *Cureus* (2023)
27. Norcliffe, A., et al.: Benchmarking Continuous Time Models for Predicting Multiple Sclerosis Progression. *Transactions on Machine Learning Research* (2023)
28. Pearl, J.: The foundations of causal inference. *Sociological Methodology* (2010)
29. Pontillo, G., et al.: Stratification of multiple sclerosis patients using unsupervised machine learning: a single-visit MRI-driven approach. *European Radiology* (2022)
30. Qian, Z., et al.: Integrating Expert ODEs into Neural ODEs: Pharmacology and Disease Progression. In: *Advances in Neural Information Processing Systems* (2021)
31. Ren, J., et al.: Balanced MSE for Imbalanced Visual Regression. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2022)
32. Rudick, R.A., et al.: Disability Progression in a Clinical Trial of Relapsing-Remitting Multiple Sclerosis: Eight-Year Follow-up. *Archives of Neurology* (11 2010)
33. Sanchez, P., et al.: Causal machine learning for healthcare and precision medicine. *Royal Society Open Science* (2022)
34. Shalit, U., et al.: Estimating individual treatment effect: generalization bounds and algorithms. *PMLR* (2017)
35. Sherstinsky, A.: Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *CoRR* (2018)
36. Sled, J., et al.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *TMI* (2002)
37. Tang, W., et al.: Soden: A scalable continuous-time survival model through ordinary differential equation networks. *The Journal of Machine Learning Research* (2022)
38. Tousignant, A., et al.: Prediction of Disease Progression in Multiple Sclerosis Patients using Deep Learning Analysis of MRI Data. In: *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*. *PMLR* (2019)
39. Verhelst, T., et al.: Uplift vs. predictive modeling: a theoretical analysis. *arXiv* (2023)
40. Vollmer, T.L., et al.: A randomized placebo-controlled phase III trial of oral laquinimod for multiple sclerosis. *J Neurology* (2014)
41. Wenao, M., et al.: Treatment Outcome Prediction for Intracerebral Hemorrhage via Generative Prognostic Model with Imaging and Tabular Data. *MICCAI* (2023)
42. Xuechen, L., et al.: Scalable Gradients for Stochastic Differential Equations. *CoRR* (2020)