Class-Balanced Deep Learning with Adaptive Vector Scaling Loss for Dementia Stage Detection

Boning Tong[1], Zhuoping Zhou[1], Davoud Ataee Tarzanagh[1], Bojian Hou[1], Andrew J. Saykin[2], Jason Moore[3], Marylyn Ritchie[1], and Li Shen[1]⋆

[1] University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Indiana University, Indianapolis, IN 46202, USA
[3] Cedars-Sinai Medical Center, Los Angels, CA 90069, USA

Abstract. Alzheimer's disease (AD) leads to irreversible cognitive decline, with Mild Cognitive Impairment (MCI) as its prodromal stage. Early detection of AD and related dementia is crucial for timely treatment and slowing disease progression. However, classifying cognitive normal (CN), MCI, and AD subjects using machine learning models faces class imbalance, necessitating the use of balanced accuracy as a suitable metric. To enhance model performance and balanced accuracy, we introduce a novel method called VS-Opt-Net. This approach incorporates the recently developed vector-scaling (VS) loss into a machine learning pipeline named STREAMLINE. Moreover, it employs Bayesian optimization for hyperparameter learning of both the model and loss function. VS-Opt-Net not only amplifies the contribution of minority examples in proportion to the imbalance level but also addresses the challenge of generalization in training deep networks. In our empirical study, we use MRI-based brain regional measurements as features to conduct the CN vs MCI and AD vs MCI binary classifications. We compare the balanced accuracy of our model with other machine learning models and deep neural network loss functions that also employ classbalanced strategies. Our findings demonstrate that after hyperparameter optimization, the deep neural network using the VS loss function substantially improves balanced accuracy. It also surpasses other models in performance on the AD dataset. Moreover, our feature importance analysis highlights VS-Opt-Net's ability to elucidate biomarker differences across dementia stages.

Keywords: Class-Balanced Deep Learning · Hyperparameter Optimization · Neuroimaging · Mild Cognitive Impairment · Alzheimer's Disease.
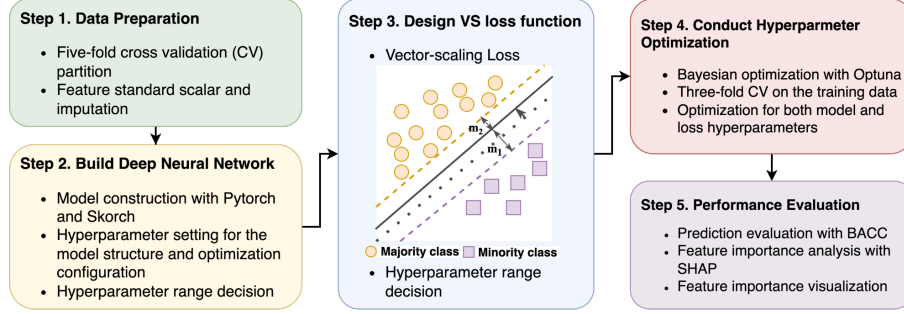
Fig. 1. VS-Opt-Net integrates the VS loss [10] into the STREAMLINE [23] pipeline and employs Bayesian optimization to adaptively learn hyperparameters for both the model and loss function. In Step 3, the VS enlarges the margin of the minority class ($\mathbf{m}_1$) relative to the majority class's margin ($\mathbf{m}_2$).

1   Introduction

Alzheimer's disease (AD) is a degenerative neurological disorder, ranked as the fifth-leading cause of death among Americans aged 65 and older [2]. It leads to irreversible cognitive decline, characterized by gradual cognitive and behavioral impairments [24]. Mild Cognitive Impairment (MCI) is a significant precursor to AD, emphasizing the need for early detection for prompt treatment and disease management [18]. However, distinguishing MCI from cognitively normal (CN) or AD subjects is challenging due to subtle brain changes observed in MCI.

Numerous machine learning algorithms excel in detecting MCI [9,11,20]. However, health datasets, including MCI detection, commonly face imbalanced class distribution [8]. For instance, the MRI data set in Alzheimer's Disease Neuroimaging Initiative (ADNI) [26,27] contains approximately twice as many MCI subjects as CN or AD subjects. Class imbalance can lead to the underrepresentation of minorities, even with highly accurate models. Techniques like data resampling, data augmentation, and class re-weighting have been used to address class imbalance in MCI classification tasks [6,8,16,17,19,29]. However, these approaches may not be as effective for overparameterized models, such as deep neural networks (DNNs), which can suffer from poor generalization [3,10,14,21]. Consequently, such models may overfit the training data, leading to discrepancies in performance when applied to unseen test data.

In light of the challenges faced by existing AD-related classification methods in overparameterized models, we present a novel Bayesian framework that achieves informative predictions for imbalanced data and minimizes generalization error. Our contributions can be summarized as follows:

• **A New Method: VS-Opt-Net (Sec. 2).** We propose VS-Opt-Net, which integrates the vector-scaling loss [10] into the STREAMLINE machine learning pipeline [22,23,25]. Utilizing Bayesian optimization, we adaptively learn hyperpa-

rameters for both the model and loss function. VS-Opt-Net not only enhances the contribution of minority examples in proportion to the imbalance level but also addresses the challenge of generalization in DNNs. For a summarized overview of VS-Opt-Net, refer to Figure 1.

• **Prediction Performance Analysis (Sec. 3).** Using MRI-based brain regional measurements, we conduct CN vs MCI and AD vs MCI binary classifications, comparing the balanced accuracy with other machine learning models employing class-balanced strategies. The results demonstrate VS-Opt-Net's superiority in the AD dataset after hyperparameter optimization.

• **Feature Importance Analysis (Sec. 3).** Besides evaluating the models' classification performance, we conduct a comparative study on the features' impact on prediction. Our findings showcase VS-Opt-Net's explanatory ability in detecting biomarker differences at various dementia stages.

2   Proposed Method

In this section, we cover classification basics, outline the VS loss, explore the STREAMLINE pipeline, and introduce our method, VS-Opt-Net.

Balanced Accuracy and VS Loss. Let (X, Y) be a joint random variable following an underlying distribution $\mathcal{P}(X, Y)$, where $X \in \mathcal{X} \subset \mathbb{R}^d$ is the input, and $Y \in \mathcal{Y} = \{1, \ldots, K\}$ is the label. Suppose we have a dataset $\mathcal{S} = (\mathbf{x}_i, y_i)_{i=1}^n$ sampled i.i.d. from a distribution $\mathcal{P}$ with input space $\mathcal{X}$ and K classes. Let $f : \mathcal{X} \to \mathbb{R}^K$ be a model that outputs a distribution over classes and let $\hat{y}_f = \arg\max_{k \in [K]} f_k(\mathbf{x})$ denote the predicted output. The balanced accuracy (BACC) is the average of the class-conditional classification accuracy:

$$\text{BACC} := \frac{1}{K} \sum_{k=1}^{K} \mathbb{P}_{\mathcal{P}_k} \Big[y = \hat{y}_f(\mathbf{x}) \Big]. \tag{BACC}$$

Our approach initially focuses on the VS loss, but it can accommodate other loss functions as well. We provide a detailed description of the VS loss and refer readers to Table 1 for SOTA re-weighting methods designed for training on imbalanced data with distribution shifts. The VS loss [10] unifies multiplicative shift [28], additive shift [14], and loss re-weighting to enhance BACC. For any $(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$, it has the following form:

$$\ell_{\text{VS}}(y, f(\mathbf{x})) := -w_y \log \left(\frac{e^{l_y f(\mathbf{x})_y + \Delta_y}}{\sum_{j=1}^{k} e^{l_j f(\mathbf{x})_j + \Delta_j}} \right). \tag{VS}$$

Here, w_j represents the classical weighting term, and l_j and Δ_j are additive and multiplicative logit adjustments. We work with $K = 2$ and aim to find logit parameters (Δ_j, l_j) that optimize BACC. When variables l and Δ are completely unknown and require adaptive optimization based on the model and datasets, we refer to VS as the VS-Opt loss function. The impact of the VS loss on improving balanced accuracy is well-studied in [10,12,21].

STREAMLINE. Simple transparent end-to-end automated machine learning (STREAMLINE) [23] is a pipeline that analyzes datasets with various models through hyperparameter optimization. It serves the specific purpose of comparing performance across datasets, machine learning algorithms, and other Automated machine learning (AutoML) tools. It stands out from other AutoML tools due to its fully transparent and consistent baseline for comparison. This is achieved through a well-designed series of pipeline elements, encompassing exploratory analysis, basic data cleaning, cross-validation partitioning, data scaling and imputation, filter-based feature importance estimation, collective feature selection, ML modeling with hyperparameter optimization over 15 established algorithms, evaluation across 16 classification metrics, model feature importance estimation, statistical significance comparisons, and automatic exporting of all results, plots, a summary report, and models. These features allow for easy application to replication data and enable users to make informed decisions based on the generated results.

VS-Opt-Net: Vector Scaling Loss Optimized for Deep Networks. The following steps introduce VS-Opt-Net, a Bayesian approach for optimizing (VS) loss for DNNs. Figure 1 provides a summary of VS-Opt-Net.

● **Step 1.** We use STREAMLINE for data preprocessing, including train-test split with stratified sampling for five-fold CV. We also impute missing values and scale features to the standard normal distribution for each dataset.

● **Step 2.** We integrate feedforward DNNs and class-balanced DNNs models into STREAMLINE using `skorch`, a PyTorch integration tool with `sklearn`. For DNNs models, we search for optimal structures by setting hyperparameters, including the number of layers and units, activation function, dropout rate, batch normalization usage, as well as optimization configurations such as the function, learning rate, batch size, and epochs. The hyperparameter ranges remain consistent across different models and classification tasks.

● **Step 3.** We integrate VS loss into STREAMLINE for DNNs adaptation. We establish decision boundaries for hyperparameters (l, Δ) in VS loss, as well as for model parameters. We optimize $\tau \in [-1, 2]$ and $\gamma \in [0, 0.5]$ for SOTA losses (see, Table 1), and $l \in [-2, 2]$ and $\Delta \in [0, 1.5]$ for VS-Opt-Net.

● **Step 4.** We employ `Optuna` [1], an open-source Python library used for hyperparameter optimization and built on top of the TPE (Tree-structured Parzen Estimator) algorithm, which is a Bayesian optimization method. We conduct a three-fold CV on the training set, performing a 100-trial Bayesian sweep to optimize both model and loss hyperparameters[4]. For the existing models, we set 'class_weight' to 'None' and 'balanced' to control the use of weights.

● **Step 5.** We report BACC for evaluating imbalanced classification performance. We use SHAP (SHapley Additive exPlanations) [13] with KernelExplainer to assess feature importance across different models, and top features are visualized by the bar plots and brain region plots.

---

[4] For existing machine learning models, optimized parameters can be found in `https://github.com/UrbsLab/STREAMLINE`

Table 1. Fixed and tunable hyperparameters for parametric CE losses. N_k denotes sample numbers for class k, and $N_{\min}$ and $N_{\max}$ represent the minimum and maximum sample numbers across all classes; π_k indicates the prior probability of class k.

Loss	Additive	Multiplicative	Optimized Hyperparameter
LDAM [3]	$l = -\frac{1}{2}(N_{\min}/N_k)^{1/4}$	-	-
LA [14]	$l = \tau \log(\pi_k)$	-	τ
CDT [28]	-	$\Delta = (N_k/N_{\max})^{\gamma}$	γ
VS [10]	$l = \tau \log(\pi_k)$	$\Delta = (N_k/N_{\max})^{\gamma}$	τ, γ
l-Opt	l	-	l
Δ-Opt	-	Δ	Δ
VS-Opt	l	Δ	l, Δ

3 Experiments

Datasets. Data for this study were sourced from the ADNI database [26,27], which aims to comprehensively assess the progression of MCI and AD through a combination of serial MRI, PET, other biological markers, and clinical evaluations. Participants granted written consent, and study protocols gained approval from respective Institutional Review Boards (IRBs)[5]. We collected cross-sectional Freesurfer MRI data from the ADNI site, merging ADNI-1/GO/2 datasets. From the total 1,470 participants (365 CN, 800 MCI, and 305 AD subjects), we selected 317 regional MRI metrics as features. These encompass cortical volume (CV), white matter volume (WMV), surface area (SA), average cortical thickness (TA), and cortical thickness variability (TSD). With these MRI measures as predictors, we performed two binary classifications with noticeable class imbalance: CN vs MCI and AD vs MCI.

Baselines. In addition to the deep neural network, we have chosen five commonly used classification models from STREAMLINE to serve as baseline models. These include elastic net, logistic regression, decision tree, random forest, and support vector machine. For all six models, the weight for the k–th class is calculated as $w_k = \pi_k^{-1} = N/(K \cdot N_k)$, where N is the total number of samples, K represents the number of classes (2 for binary classification), N_k is the number of samples for class k, and π_k is the prior probability of class k. We conducted a comparison of DNNs models employing various class-balanced loss functions. Besides the traditional cross-entropy (CE) and weighted cross-entropy (wCE) losses, we evaluated our model against state-of-the-art (SOTA) losses listed in Table 1. These losses incorporate at least one logit adjustment based on class distributions. For models with LA, CDT, and VS loss, we utilized Bayesian optimization to select optimal τ and γ values, which determined

---

[5] For the latest information, visit `www.adni-info.org`.

Table 2. Comparison of BACC (mean $\pm$ std) for two binary classification tasks using wCE losses with optimized weights w_y and default weight ($w_y = 1$). The table emphasizes that re-weighting alone is ineffective for deep neural networks.

Model	CN VS MCI		AD VS MCI	
	$w_y = 1$	Optimzied w_y	$w_y = 1$	Optimzied w_y
Elastic Net	0.580±0.011	0.650±0.026	0.652±0.040	0.732±0.019
Logistic Regression	0.592±0.037	**0.657±0.042**	**0.738±0.034**	0.742±0.038
Decision Tree	0.569±0.020	0.612±0.027	0.628±0.018	0.679±0.032
Random Forest	0.555±0.018	0.639±0.015	0.657±0.023	0.724±0.024
Support Vector Machine	0.569±0.010	0.650±0.035	0.641±0.014	**0.744±0.042**
Deep Neural Network	**0.606±0.009**	0.633±0.032	0.700±0.055	0.709±0.023

Table 3. Balanced accuracy BACC for classification tasks using DNNs models with different loss functions. Cross-validation results are shown as mean±std in each cell. We tuned the τ and γ in LA, CDT, and VS losses to find l and Δ, following parameters in Table 1. For l-Opt, Δ-Opt, and VS-Opt losses, we adaptively optimized l and Δ.

Loss	CN vs MCI	AD vs MCI
CE	0.606±0.009	0.700±0.055
wCE (w_y)	0.633±0.032	0.709±0.023
LDAM (l)	0.625±0.033	0.726±0.046
LA (l)	0.611±0.037	0.733±0.028
CDT (Δ)	0.608±0.022	0.715±0.033
VS ($l + \Delta$)	0.646±0.035	0.745±0.039
l-Opt	0.641±0.029	0.738±0.037
Δ-Opt	0.608±0.017	0.727±0.043
VS-Opt	**0.669±0.048**	**0.754±0.026**

l and Δ. Additionally, we introduced two novel approaches: l-Opt and Δ-Opt loss, where we directly optimized logit adjustments l and Δ through Bayesian optimization without class distribution constraints. Furthermore, our proposed method, VS-Opt-Net, optimizes l and Δ together in the VS-Opt loss function.

Prediction Performance Results. We evaluated the prediction performance of various machine learning models for the CN vs MCI and AD vs MCI classification tasks (Table 2). BACC was used to calculate the mean and standard deviation. When comparing models with and without class-balanced weights, we observed that all models showed improvement in BACC after incorporating the weight. However, it is worth noting that the weighted deep neural network underperformed compared to the weighted logistic regression in the CN vs MCI classification, and the weighted SVM in the AD vs MCI classification.

Table 3 compares DNNs models using different class-balanced loss functions. Our numerical analysis shows that models incorporating both additive and mul-
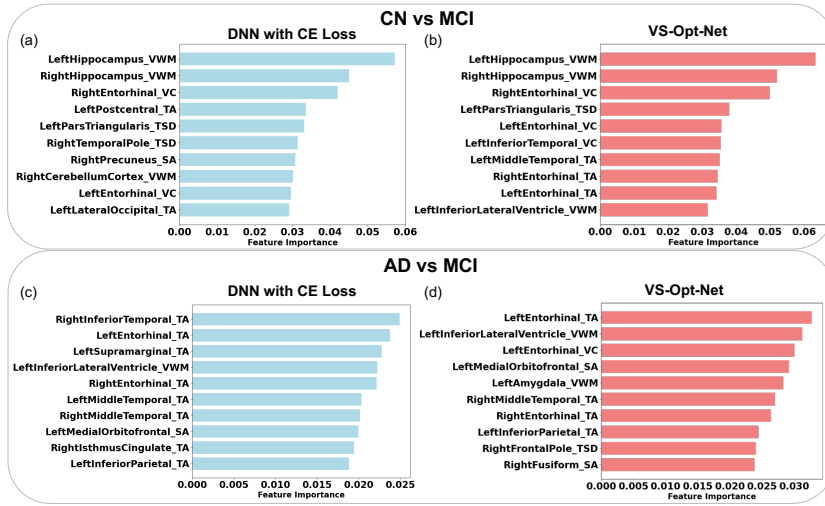
Fig. 2. SHAP feature importance for DNNs with cross-entropy loss (a,c) and VS-Opt-Net (b,d). (a-b) Top regions for CN vs MCI classification. (c-d) Top regions for AD vs MCI classification. Each figure displays the top 10 features for each case.

tiplicative logit adjustments achieve higher BACC scores than those with only one adjustment, consistent with previous findings in image recognition [10]. Additionally, directly optimizing adjustment parameters with VS-Opt leads to improved prediction performance compared to baselines, enabling our approach to outperform all baseline models.

Feature Importance and Top-Ranked Regions. We analyze feature contributions and assess model classification performance. Figure 2 depicts SHAP feature importance for DNNs with CE and VS-Opt-Net, while Fig. 3 reveals significant brain regions by volume (cortical/white matter), cortical thickness (average/standard deviation), and surface area for VS-Opt-Net. Notably, top-ranking brain regions exhibit similarity between the models, with some regions notably more influential in our model. Cortical/white matter volume and average cortical thickness hold prominent predictive power. Noteworthy features distinguishing CN and MCI encompass hippocampus and right entorhinal cortex volumes. Our model emphasizes the volume of the left entorhinal and left inferior temporal gyri, along with the average thickness of the left middle temporal gyrus—features given less priority by traditional DNNs. For AD vs MCI, key contributors are average thickness of the left entorhinal and volume of the left inferior lateral ventricle. Additionally, contributions from the left entorhinal area and amygdala volumes increase.

The volume reductions of the entorhinal cortex and hippocampus are biomarkers of early Alzheimer's disease. According to prior studies, CN and MCI can be differentiated more accurately using hippocampal volume than lateral neocortical measures [4], which aligns with our feature importance analysis. Additionally,
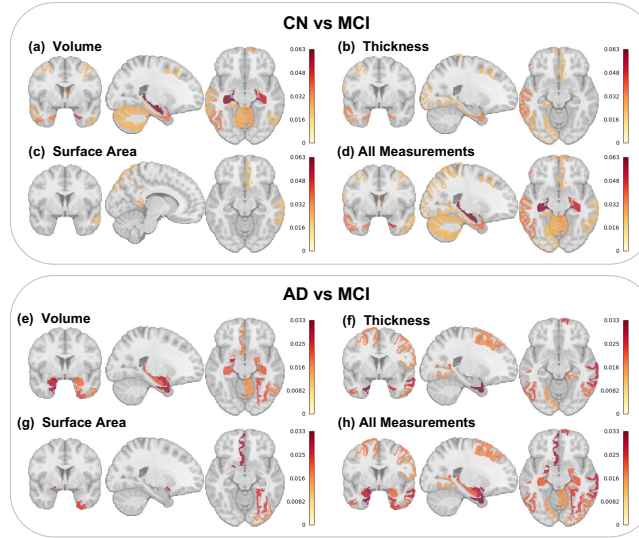
Fig. 3. Brain visualization of the leading 40 features for VS-Opt-Net. Colormap indicates SHAP feature importance; darker shades signify higher significance. Panels (a-d) reveal top features for CN vs MCI classification, while panels (e-h) showcase prime features for AD vs MCI classification. Notably, (a-c) and (e-g) spotlight regions of heightened importance in terms of volume, thickness, and surface area measures for both prediction categories. Panel (d) consolidates (a-c), while panel (h) amalgamates (e-g), displaying the highest importance value when a region encompasses multiple measurements.

studies have found a significant brain atrophy and thickness decrease in the inferior and middle temporal gyri for MCI patients compared with healthy control [7]. Other studies have reported that AD vs MCI identification is improved by using the entorhinal cortex rather than the hippocampus [5] and the outward deformation of the lateral ventricles. Besides, there is a significant atrophy for the left amygdala when comparing MCI and AD subjects, which is related to the AD severity [15]. The above findings demonstrate the explanatory ability for our model to differentiate between different stages of dementia.

4   Conclusion

We introduced VS-Opt-Net, a novel model integrating the VS loss into STREAM-LINE with Bayesian optimization for hyperparameter tuning. It effectively addressed class imbalance and generalization challenges by enhancing the contribution of minority examples. In binary classifications of CN vs MCI and AD vs MCI using MRI-based brain regional measurements, VS-Opt-Net significantly improved BACC, outperforming other models in the AD dataset. Our feature importance analysis revealed successful biomarker explanation at different dementia stages.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
2. Association, A., et al.: 2012 alzheimer's disease facts and figures. Alzheimer's & Dementia **8**(2), 131–168 (2012)
3. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning imbalanced datasets with label-distribution-aware margin loss. Advances in neural information processing systems **32** (2019)
4. De Santi, S., de Leon, M.J., Rusinek, H., Convit, A., Tarshish, C.Y., Roche, A., Tsui, W.H., Kandil, E., Boppana, M., Daisley, K., et al.: Hippocampal formation glucose metabolism and volume losses in mci and ad. Neurobiology of aging **22**(4), 529–539 (2001)
5. Du, A., Schuff, N., Amend, D., Laakso, M., Hsu, Y., Jagust, W., Yaffe, K., Kramer, J., Reed, B., Norman, D., et al.: Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and alzheimer's disease. Journal of Neurology, Neurosurgery & Psychiatry **71**(4), 441–447 (2001)
6. Dubey, R., Zhou, J., Wang, Y., Thompson, P.M., Ye, J., Initiative, A.D.N., et al.: Analysis of sampling techniques for imbalanced data: An n= 648 adni study. NeuroImage **87**, 220–241 (2014)
7. Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., Initiative, A.D.N., et al.: Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage **39**(4), 1731–1743 (2008)
8. Hu, S., Yu, W., Chen, Z., Wang, S.: Medical image reconstruction using generative adversarial network for alzheimer disease assessment with class-imbalance problem. In: 2020 IEEE 6th international conference on computer and communications (ICCC). pp. 1323–1327. IEEE (2020)
9. Kim, D., Kim, S., Risacher, S.L., Shen, L., Ritchie, M.D., Weiner, M.W., Saykin, A.J., Nho, K., the Alzheimer's Disease Neuroimaging, I.: A graph-based integration of multimodal brain imaging data for the detection of early mild cognitive impairment (e-mci). Multimodal Brain Image Anal (2013) **8159**, 159–169 (2013)
10. Kini, G.R., Paraskevas, O., Oymak, S., Thrampoulidis, C.: Label-imbalanced and group-sensitive classification under overparameterization. Advances in Neural Information Processing Systems **34**, 18970–18983 (2021)
11. Li, J., Bian, C., Chen, D., Meng, X., Luo, H., Liang, H., Shen, L.: Persistent feature analysis of multimodal brain networks using generalized fused lasso for emci identification. Med Image Comput Comput Assist Interv **12267**, 44–52 (2020)
12. Li, M., Zhang, X., Thrampoulidis, C., Chen, J., Oymak, S.: Autobalance: Optimized loss functions for imbalanced data. Advances in Neural Information Processing Systems **34**, 3163–3177 (2021)
13. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)
14. Menon, A.K., Jayasumana, S., Rawat, A.S., Jain, H., Veit, A., Kumar, S.: Long-tail learning via logit adjustment. arXiv preprint arXiv:2007.07314 (2020)
15. Miller, M.I., Younes, L., Ratnanather, J.T., Brown, T., Reigel, T., Trinh, H., Tang, X., Barker, P., Mori, S., Albert, M.: Amygdala atrophy in mci/alzheimer's disease

in the biocard cohort based on diffeomorphic morphometry. In: Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention. vol. 2012, p. 155. NIH Public Access (2012)

16. Oktavian, M.W., Yudistira, N., Ridok, A.: Classification of alzheimer's disease using the convolutional neural network (cnn) with transfer learning and weighted loss. arXiv preprint arXiv:2207.01584 (2022)

17. Puspaningrum, E.Y., Wahid, R.R., Amaliyah, R.P., et al.: Alzheimer's disease stage classification using deep convolutional neural networks on oversampled imbalance data. In: 2020 6th Information Technology International Seminar (ITIS). pp. 57–62. IEEE (2020)

18. Rasmussen, J., Langerman, H.: Alzheimer's disease–why we need early diagnosis. Degenerative neurological and neuromuscular disease pp. 123–130 (2019)

19. Sadegh-Zadeh, S.A., Fakhri, E., Bahrami, M., Bagheri, E., Khamsehashari, R., Noroozian, M., Hajiyavand, A.M.: An approach toward artificial intelligence alzheimer's disease diagnosis using brain signals. diagnostics **13**(3), 477 (2023)

20. Shen, L., Kim, S., Qi, Y., Inlow, M., Swaminathan, S., Nho, K., Wan, J., Risacher, S.L., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Saykin, A.J., ADNI: Identifying neuroimaging and proteomic biomarkers for MCI and AD via the elastic net. Multimodal Brain Image Analysis **7012**, 27–34 (2011)

21. Tarzanagh, D.A., Hou, B., Tong, B., Long, Q., Shen, L.: Fairness-aware class imbalanced learning on multiple subgroups. In: Uncertainty in Artificial Intelligence. pp. 2123–2133. PMLR (2023)

22. Tong, B., Risacher, S.L., Bao, J., Feng, Y., Wang, X., Ritchie, M.D., Moore, J.H., Urbanowicz, R., Saykin, A.J., Shen, L.: Comparing amyloid imaging normalization strategies for alzheimer's disease classification using an automated machine learning pipeline. AMIA Jt Summits Transl Sci Proc **2023**, 525–533 (2023)

23. Urbanowicz, R., Zhang, R., Cui, Y., Suri, P.: Streamline: A simple, transparent, end-to-end automated machine learning pipeline facilitating data analysis and algorithm comparison. In: Genetic Programming Theory and Practice XIX, pp. 201–231. Springer (2023)

24. Uwishema, O., Mahmoud, A., Sun, J., Correia, I.F.S., Bejjani, N., Alwan, M., Nicholas, A., Oluyemisi, A., Dost, B.: Is alzheimer's disease an infectious neurological disease? a review of the literature. Brain and Behavior **12**(8), e2728 (2022)

25. Wang, X., Feng, Y., Tong, B., Bao, J., Ritchie, M.D., Saykin, A.J., Moore, J.H., Urbanowicz, R., Shen, L.: Exploring automated machine learning for cognitive outcome prediction from multimodal brain imaging using streamline. AMIA Jt Summits Transl Sci Proc **2023**, 544–553 (2023)

26. Weiner, M.W., Veitch, D.P., Aisen, P.S., et al.: The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement **9**(5), e111–94 (2013)

27. Weiner, M.W., Veitch, D.P., Aisen, P.S., et al.: Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. Alzheimer's & Dementia **13**(4), e1–e85 (2017)

28. Ye, H.J., Chen, H.Y., Zhan, D.C., Chao, W.L.: Identifying and compensating for feature deviation in imbalanced deep learning. arXiv preprint arXiv:2001.01385 (2020)

29. Zeng, L., Li, H., Xiao, T., Shen, F., Zhong, Z.: Graph convolutional network with sample and feature weights for alzheimer's disease diagnosis. Information Processing & Management **59**(4), 102952 (2022)