



HAL
open science

Semi-supervised Domain Adaptation for Automatic Quality Control of FLAIR MRIs in a Clinical Data Warehouse

Sophie Loizillon, Olivier Colliot, Lydia Chougar, Sebastian Stroer, Yannick Jacob, Aurélien Maire, Didier Dormont, Ninon Burgos

► **To cite this version:**

Sophie Loizillon, Olivier Colliot, Lydia Chougar, Sebastian Stroer, Yannick Jacob, et al.. Semi-supervised Domain Adaptation for Automatic Quality Control of FLAIR MRIs in a Clinical Data Warehouse. DART 2023 - 5th MICCAI Workshop on Domain Adaptation and Representation Transfer, Oct 2023, Vancouver (BC), Canada. pp.84-93, 10.1007/978-3-031-45857-6_9 . hal-04273997

HAL Id: hal-04273997

<https://hal.science/hal-04273997>

Submitted on 7 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-supervised Domain Adaptation for Automatic Quality Control of FLAIR MRIs in a Clinical Data Warehouse

Sophie Loizillon¹, Olivier Colliot¹, Lydia Chougar¹ Sebastian Stroer³ Yannick Jacob² Aurélien Maire², Didier Dormont^{1,3}, and Ninon Burgos¹

¹ Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

² AP-HP, Innovation & Données – Département des Services Numériques, Paris, France

³ AP-HP, Hôpital Pitié-Salpêtrière, DMU DIAMENT, Dep. of Neuroradiology, Paris, France

sophie.loizillon@gmail.com

Abstract. Domain adaptation is a very useful approach to exploit the potential of clinical data warehouses, which gather a vast amount of medical imaging encompassing various modalities, sequences, manufacturers and machines. In this study, we propose a semi-supervised domain adaptation (SSDA) framework for automatically detecting poor quality FLAIR MRIs within a clinical data warehouse. Leveraging a limited number of labelled FLAIR and a large number of labelled T1-weighted MRIs, we introduce a novel architecture based on the well known Domain Adversarial Neural Network (DANN) that incorporates a specific classifier for the target domain. Our method effectively addresses the covariate shift and class distribution shift between T1-weighted and FLAIR MRIs, surpassing existing SSDA approaches by more than 10 percent points.

Keywords: Domain Adaptation · Deep Learning · Magnetic Resonance Imaging · Clinical Data Warehouse.

1 Introduction

Machine learning algorithms aim to learn a model from training samples by minimising a cost function. However, for these models to work optimally, it is crucial that the training data and the test data share similar distributions. When the distribution of the training dataset differs from that of the test data, the performance of the model degrades. This would be the case when training an algorithm to rate the quality of T1-weighted (T1w) magnetic resonance (MR) images but applying it to another MRI sequence, e.g. fluid attenuated inversion recovery (FLAIR). To overcome this problem, various approaches have been proposed in the field of domain adaptation [1,2,3,4]. These approaches aim to develop robust models that can effectively adapt to various data distributions that differ from those on which they have been trained.

Domain adaptation has first been tackled in the literature in an unsupervised manner [1,5], with the objective of reducing the discrepancy between the source (e.g., T1w) and target (e.g., FLAIR) domains relying on labelled data from the source domain only. This is commonly referred to as unsupervised domain adaptation (UDA). Adversarial learning has achieved promising results by extracting domain invariant features. Adversarial learning attempts to extract relevant features to fool the domain classifier which aims to differentiate the source and target domains [5]. An effective adversarial learning approach is that of the Domain Adversarial Neural Network (DANN), which is composed of a shared feature extractor and a domain discriminator [1]. The domain discriminator aims to distinguish the source and target samples. At the same time, the feature extractor is trained to fool the domain discriminator in order to learn domain invariant features. While this method has proved its worth in recent years, Zhao et al. [6] warned that learning domain-independent features does not guarantee good generalisation to the target domain, particularly when the class distribution between domains differs.

In some cases, a few labelled samples are available from the target domain, which can help in improving the performance of the model. Saito et al. [2] introduced the notion of semi-supervised domain adaptation (SSDA), a variant of UDA where a limited number of labelled target samples is available. Specifically, they proposed a method called minimax entropy (MME). With the use of the mini-max paradigm, the presence of a small number of labelled target samples can considerably improve the performance of CNN models. The method consists of first estimating domain-invariant prototypes by maximising the entropy, and then clustering the features around the estimated prototypes, this time by minimising the same entropy. Several approaches focus on reducing the intra-domain divergence in the target domain [3,4]. Jiang et al. [3] designed an effective Bidirectional Adversarial Training (BiAT) to attract unaligned sub-distributions from the target domain to the corresponding source sub-distributions. While many existing SSDA approaches rely on adversarial techniques, there has been a recent interest in the use of contrastive learning. These methods, that show impressive success in self-supervised learning, aim to acquire an embedding space where similar pairs of samples are grouped together, while dissimilar pairs are pushed apart. In the context of domain adaptation, a common approach is to consider as similar pairs unlabelled target images and their corresponding augmented versions. On the other hand, dissimilar pairs are formed by combining unlabelled target images with labelled source images [7,8].

Due to the large variety of modalities and sequences, the field of medical imaging has turned its attention to the question of UDA and SSDA. Adversarial learning has been applied to various medical image classification and segmentation tasks [9,10,11]. An extension of the DANN to an encoder-decoder architecture is proposed in [9] and used for segmentation purposes in volume electron microscopy imaging. The network was initially trained using UDA techniques and subsequently fine-tuned on the target domain. A comparative study of the original DANN, a semi-supervised DANN, and fine-tuning techniques for white

matter lesion segmentation is presented in [10]. Best results were obtained with the semi-supervised DANN. However, it was trained by passing all source and target samples through the same label classifier, which does not seem ideal especially when there is a distribution shift between source and target. Contrastive learning has also been used for domain adaptation in medical image analysis, for instance to reduce the gap between adult (source domain) and children (target domain) chest X-rays for automatic pneumonia diagnosis [11].

A sub-field of medical imaging where domain adaptation would be particularly necessary is that focusing on clinical routine data, such as those gathered in clinical data warehouses (CDWs). Having access to extensive and diverse clinical imaging data from multiple centres present an exceptional opportunity for the development of computational tools. To efficiently exploit CDWs, a model initially developed for a specific modality or sequence should have the ability to generalise to other modalities or sequences without the need to manually re-annotate large amounts of data. For example, recent studies have emphasised the importance of conducting automatic quality control (QC) on imaging data before deploying machine learning algorithms on CDWs [12]. To address this need, a framework was previously developed for the automatic QC of T1w brain MRIs. However, this framework, specific to the T1w sequence, required significant manual annotation efforts involving 5500 image annotations [13]. Our aim is to expand this framework to encompass FLAIR MRIs, while minimising the need for extensive new manual annotations using domain adaptation.

In this paper, we introduce a novel SSDA method for the automatic detection of poor quality FLAIR MRIs in a CDW while using a limited amount of labelled target data. In particular, we propose to incorporate a specific classifier for the target data into the well-known DANN [1]. We conduct a comparative analysis between our proposed method and the state-of-the-art SSDA architectures [2,10,14].

2 Methods

2.1 Dataset description

This work is built upon a large clinical routine dataset containing 3D T1w and FLAIR brain MR images of adult patients scanned in hospitals of the Greater Paris area (Assistance Publique-Hôpitaux de Paris [AP-HP]). The dataset is composed of 5500 T1w and 5858 FLAIR MRIs acquired on respectively 30 and 22 different scanners from four manufacturers (Siemens, GE, Philips, and Toshiba) which had been randomly selected from the CDW [13].

5500 T1w and 858 FLAIR MRIs were manually annotated with respect to the MRI quality. A score was given by two annotators to evaluate the noise, contrast and motion within the image on a three-point scale (0: no artefact, 1: moderate artefact, 2: severe artefact) [13]. In case of disagreement, the consensus score was chosen as the most severe of the two for T1w MRIs [13], while for the FLAIR images, both annotators reviewed the problematic images and agreed on a consensus, given that the number of manually annotated images was limited.

Based on the scores assigned to the motion, contrast and noise characteristics, we defined different tiers: *tier 1* represents good quality images (score of 0 for all characteristics), *tier 2* stands for medium quality images (score 1 for at least one characteristic and no score 2), and *tier 3* for bad quality images (score 2 for at least one characteristic). If the images did not contain full 3D T1w or FLAIR MRI of the whole brain, such as segmented tissue images or truncated images, they were classified as straight reject (*SR*). Figure 1 shows the distribution of the different tiers according to the modality in the form of a Sankey plot.

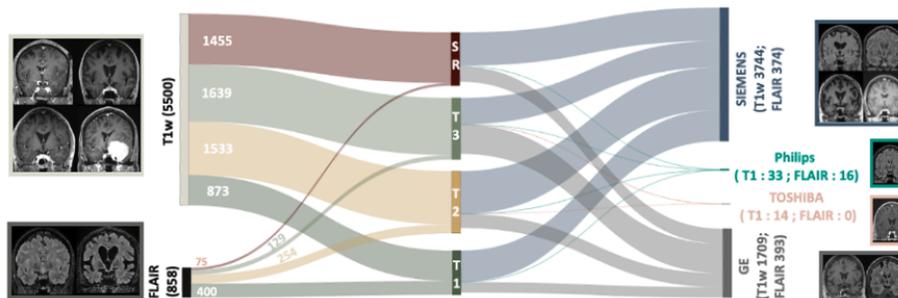


Fig. 1. Sankey plot analysis of labelled T1w and FLAIR MRIs highlighting the tier distribution for both sequences and the distribution of manufacturers across tiers.

2.2 Image pre-processing

The T1w and FLAIR MRIs were pre-processed using Clinica and its `{t1|flair}-linear` pipeline [15] that relies on ANTs [16]. First, bias field correction was applied using the N4ITK method. Next, an affine registration to the MNI space was performed using a specific template for each of the T1w and FLAIR sequences. The images were then cropped to remove the background, resulting in $169 \times 208 \times 179$ images and 1 mm isotropic voxels.

2.3 Proposed approach

We developed a semi-supervised approach based on the unsupervised DANN method to detect poor quality (i.e. *tier 3*) FLAIR in a clinical data warehouse.

The DANN [1], which aims to learn domain invariant features, was originally designed for unsupervised learning. The network is composed of a shared feature extractor (F) and two classifiers: a domain classifier (C_{dom}) and a label predictor classifier. The goal of this architecture is to minimise the *label prediction loss* for accurate label prediction on the source domain and maximise the *domain confusion loss* to align the feature distributions of the source and target domains. Zhao et al. [6] showed that learning domain invariant features

does not necessarily guarantee the generalisation of the model to the target domain, in particular when the class distributions change between the source and target domains. We adapted the DANN architecture to the context of SSDA by adding a target label predictor classifier C_T , as shown in Fig 2. Compared with the approach of Sundaresan et al. [10], where all source and target data pass through the same label classifier, the addition of a specific classifier to the target data aims at handling the class distribution shift between the *tier 1/2* and *tier 3* classes for the T1w and FLAIR MRIs. While for the T1w sequence, the two classes are slightly unbalanced (*tier 1/2*: 59% and *tier 3*: 41%), the imbalance is much larger for the FLAIR (*tier 1/2*: 84% and *tier 3*: 16%).

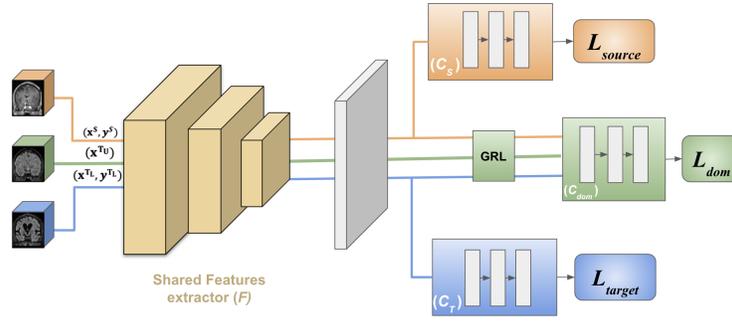


Fig. 2. Semi Supervised adapted DANN architecture composed of a domain invariant feature extractor (F), a source label classifier (C_S), a target label classifier (C_T) and a domain classifier (C_{dom}). A gradient reverse layer (GRL) multiplies the gradient by a negative value when backpropagating to maximise the loss of the domain discriminator.

We denote the source dataset, consisting of labelled T1w MRIs, as $D_S = (x_i^S, y_i^S)_{i=1}^{N_S}$. For the target domain, we have two datasets: the labelled target samples $D_{T_L} = (x_i^{T_L}, y_i^{T_L})_{i=1}^{N_{T_L}}$ and the unlabelled target samples $D_{T_U} = (x_i^{T_U})_{i=1}^{N_{T_U}}$. We will refer to the labelled data from the two domains using the following notation $D_L = D_S \cup D_{T_L}$. The overall loss for the semi supervised DANN is

$$L = \underbrace{L_{pred}(F, C_S, C_T, D_L)}_{\text{Label Prediction Loss}} - \lambda \cdot \underbrace{L_{dom}(F, C_{dom}, D_L, D_{T_U}, d)}_{\text{Domain Confusion Loss}} \quad (1a)$$

where

$$L_{pred} = - \underbrace{\sum_{k=1}^K (y_i^S)_k \log(C_S(F(x_i^S)))_k}_{\text{Source Label Prediction}} - \underbrace{\sum_{k=1}^K (y_i^{T_L})_k \log(C_T(F(x_i^{T_L})))_k}_{\text{Target Label Prediction}} \quad (1b)$$

$$L_{dom} = d_i \log \frac{1}{C_{dom}(F(x_i))} + (1 - d_i) \log \frac{1}{1 - C_{dom}(F(x_i))} \quad (1c)$$

with d_i the domain label of the i -th sample, which indicates whether it is a T1w (source domain) or a FLAIR (target domain), and $K = 2$ is the number of classes for the source and target label classifier (i.e., task *tier1/2* vs *tier3*).

As in the classical DANN [1], the model is initialised using a pre-trained model without the domain classifier. This model is usually trained on the source domain that has been labelled. Here we pre-trained the network using all the labelled data from the source and target domains. The training procedure is detailed in Algorithm 1. We structured each iteration to include three mini-batches. A first mini-batch contained labelled samples from the source domain (b_S), another consisted of labelled samples from the target domain (b_{T_L}), and the final mini-batch comprised unlabelled samples from the target domain (b_{T_U}). This arrangement ensured that a target labelled data sample was present in every batch, effectively influencing the training process.

Algorithm 1 Proposed SSDA Algorithm

Require:

- 1: Source samples $D_S = \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$
 - 2: Target labelled samples $D_{T_L} = \{(x_i^{T_L}, y_i^{T_L})\}_{i=1}^{N_{T_L}}$
 - 3: Target unlabelled samples $D_{T_U} = \{(x_i^{T_U})\}_{i=1}^{N_{T_U}}$
 - 4: Hyperparameters : λ , learning rate η
 - 5: Pre-trained model with all labelled samples (D_S, D_{T_L}) and $\lambda = 0$
 - 6: **for** each epoch **do**
 - 7: **for** each b_S, b_{T_L}, b_{T_U} **do**
 - 8: **# Forward pass**
 - 9: $\hat{y}^S \leftarrow C_S(F(b_S, \theta_F), \theta_{C_S})$
 - 10: $\hat{y}^T \leftarrow C_T(F(b_{T_L}, \theta_F), \theta_{C_T})$
 - 11: $\hat{d} \leftarrow C_{\text{dom}}(F(b_S + b_{T_L} + b_{T_U}, \theta_F), \theta_{C_{\text{dom}}})$
 - 12: **# Loss computation**
 - 13: $L \leftarrow L_S(\hat{y}^S, y^S) + L_{T_L}(\hat{y}^{T_L}, y^{T_L}) + \lambda L_{\text{dom}}(\hat{d}, d)$
 - 14: **# Backward pass & model weight update**
 - 15: $\theta_F \leftarrow \theta_F - \eta \times \nabla_{\theta_F} L$
 - 16: $\theta_{C_S} \leftarrow \theta_{C_S} - \eta \times \nabla_{\theta_{C_S}} L$
 - 17: $\theta_{C_T} \leftarrow \theta_{C_T} - \eta \times \nabla_{\theta_{C_T}} L$
 - 18: **end for**
 - 19: **end for**
-

2.4 Experiments

We aimed to assess the ability of the proposed and different existing SSDA approaches to detect bad quality FLAIR MRIs, which corresponds to the classification task *tier 1/2* vs *tier 3*. Before starting the experiments, we excluded all the *SR* images that contained MRIs which were not full 3D MRIs of the whole brain (1455 T1w and 75 FLAIR), e.g., truncated images or segmented brain tissue images. We built the FLAIR MRIs test set by randomly selecting 480 manually

annotated images while ensuring they shared the same distribution of tiers, manufacturers, and field strengths as the images in the training/validation set. The remaining 303 labelled FLAIR MRIs were split into training and validation using a 5-fold cross validation. In a similar manner, the 5000 unlabelled FLAIR MRIs were divided into training and validation sets, taking into account the same distribution of manufacturers and field strengths. As for the T1w dataset composed of 3660 labelled samples, we employed the identical split as Bottani et al. [13], where the images were split between training and validation using a 5-fold cross validation respecting the same tiers distribution. We conducted a comparative study between our proposed approach and three well-known SSDA methods that use labelled data from the source and target domains to jointly train a network, and unlabelled samples to regularise it: a semi-supervised DANN without target label classifier [10], mini-max entropy (MME) [2] and entropy minimisation (ENT) [14].

MME [2] extracts discriminating and domain-invariant features using unlabelled target data by estimating domain-invariant prototypes. Domain adaptation is performed by first maximising the entropy of unlabelled target data according to the classifier to estimate prototypes that are invariant to the domain. Then, the entropy is minimised regarding to the feature extractor to make target features well-clustered around the prototypes.

ENT [14] relies on entropy minimisation using labelled source and target data, along with unlabelled target data. The entropy is computed on the unlabelled target samples, and the network is trained to minimise this entropy. In contrast to MME, ENT does not involve a maximisation process.

All the experiments were conducted using the ClinicaDL software, implemented in PyTorch [17]. The Conv5FC3 network, consisting of five convolutional neural network layers and three fully connected layers, was used for every experiment as its performance was as good as that of more sophisticated CNN architectures (Inception, ResNet) on the T1w MRI QC task [13].

3 Results

The main results for the detection of bad quality MRIs (*tier 3*) within the CDW are shown in Table 1. For each method, we report the results on the independent T1w and FLAIR test sets. We first trained the baseline for the source domain (Baseline T1w) using the 3660 labelled T1w MRIs. The model exhibited excellent balanced accuracy (BA) on the T1w test set (83.51%) but the performance on the FLAIR test set dramatically dropped (50.06%). We performed the same experiments training on the target domain using the 303 labelled FLAIR MRIs (Baseline FLAIR). We obtained a low BA of 68.23% on the FLAIR test set. Finally, the baseline trained with all the labelled data (Baseline T1 + FLAIR) achieved an excellent BA on the T1w but a low BA on FLAIR. With regard to the results obtained for the different SSDA methods, the approach we have proposed outperforms the DANN, MME and ENT by 10, 13 and 14 percent points respectively on FLAIR MRIs, reaching a BA of 76.81%. The evaluation

of the different models, including specificity and sensitivity metrics, is presented in Table 1 of the supplementary material.

Table 1. Results for the detection of bad quality images (*tier 3*) within the T1w and FLAIR test sets from the CDW. We report the mean and empirical standard deviation across the five folds of the balanced accuracy (BA), which is defined as the mean of the specificity and sensitivity. For Manual Annotation, the BA corresponds to the average BA of the two annotators with the consensus.

Approaches	N° of	N° of	N° of	T1w	FLAIR
	S_i	T_l	T_u	BA	BA
Manual Annotation	3660	303	–	91.56	86.54
Baseline T1w	3660	0	0	83.51 ± 0.93	50.06 ± 1.71
Baseline FLAIR	0	303	0	48.20 ± 0.81	69.90 ± 2.38
Baseline T1w + FLAIR	3660	303	0	82.37 ± 0.85	58.05 ± 4.45
DANN [10]	3660	303	5000	81.97 ± 1.39	66.91 ± 2.37
ENT [14]	3660	303	5000	80.07 ± 1.99	62.21 ± 4.45
MME [2]	3660	303	5000	77.39 ± 3.82	63.31 ± 4.16
Proposed Approach	3660	303	5000	80.59 ± 1.62	76.81 ± 0.68

The observed results of the T1w baseline are in line with expectations, considering the training was conducted only on T1w images. These T1w images possess different intensity distributions when compared to the target distribution of FLAIR MRIs ($P_S(X, Y) \neq P_T(X, Y)$), which consequently led to a decline in performance when applied to FLAIR samples. Similarly, the FLAIR baseline demonstrated poor results on T1w MRIs due to the covariate shift between the two sequences. Furthermore, the limited amount of manually annotated FLAIR images (303 samples) contributed to a low balanced accuracy (BA: 68%) on the FLAIR test set. Finally, the T1w+FLAIR baseline exhibited satisfying results on the T1w test set. However, since the training labels are highly dominated by T1w images, it led to a low BA on the FLAIR test set. This underlines the significance of employing SSDA techniques to develop models that are robust to different MRI sequences.

The poor results obtained from the DANN, MME and ENT on the target domain (BA < 67%) underline the fact that learning domain invariant features is not sufficient in particular in the case of class distribution shift [6]. Indeed, while the distribution between labels was slightly unbalanced (2406 vs 1639 images) for the source domain, it was extremely unbalanced for the target domain (654 vs 129). This difference in class distributions between the T1w (source) and FLAIR (target) MRIs explains the poor performance of these methods. With our proposed approach of incorporating a second label classifier dedicated to the target modality, we were able to overcome these limitations and achieved an important improvement of more than 10 percent points. Thus, this model will be applicable to filter and select good quality FLAIR MRIs from the AP-HP CDW, enabling users to take advantage of these sequences for conducting further studies.

4 Conclusion

In this paper, we propose a new SSDA architecture based on the original DANN [1] for the detection of bad quality FLAIR MRIs in a CDW. By adding a second label classifier specifically for the target domain, we were able to overcome the covariate shift and the class distributions shift between the source and target domains. We achieved a balanced accuracy of 76.8% on the FLAIR test set and outperforms the DANN, MME and ENT by 10, 13 and 14 percent points respectively.

References

1. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. *Journal of Machine Learning Research* **17**(1) (2016) 2096–2030
2. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (2019) 8050–8058
3. Jiang, P., Wu, A., Han, Y., Shao, Y., Qi, M., Li, B.: Bidirectional Adversarial Training for Semi-Supervised Domain Adaptation. In: *IJCAI*. (2020) 934–940
4. Kim, T., Kim, C.: Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In: *Computer Vision–ECCV 2020*, Springer (2020) 591–607
5. HassanPour Zonoozi, M., Seydi, V.: A Survey on Adversarial Domain Adaptation. *Neural Processing Letters* (2022) 1–41
6. Zhao, H., Des Combes, R.T., Zhang, K., Gordon, G.: On learning invariant representations for domain adaptation. In: *International Conference on Machine Learning*, PMLR (2019) 7523–7532
7. Singh, A.: Clda: Contrastive learning for semi-supervised domain adaptation. *Advances in Neural Information Processing Systems* **34** (2021) 5089–5101
8. Thota, M., Leontidis, G.: Contrastive domain adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2021) 2209–2218
9. Roels, J., Hennies, J., Saeys, Y., Philips, W., Kreshuk, A.: Domain adaptive segmentation in volume electron microscopy imaging. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE (2019) 1519–1522
10. Sundaresan, V., Zamboni, G., Dinsdale, N.K., Rothwell, P.M., Griffanti, L., Jenkinson, M.: Comparison of domain adaptation techniques for white matter hyperintensity segmentation in brain MR images. *Medical Image Analysis* **74** (2021) 102215
11. Feng, Y., Wang, Z., Xu, X., Wang, Y., Fu, H., Li, S., Zhen, L., Lei, X., Cui, Y., Ting, J.S.Z.: Contrastive domain adaptation with consistency match for automated pneumonia diagnosis. *Medical Image Analysis* **83** (2023) 102664
12. Bottani, S., Burgos, N., Maire, A., Saracino, D., Ströer, S., Dormont, D., Colliot, O.: Evaluation of MRI-based machine learning approaches for computer-aided diagnosis of dementia in a clinical data warehouse. Preprint: <https://hal.science/hal-03656136> (2023)

13. Bottani, S., Burgos, N., Maire, A., Wild, A., Ströer, S., Dormont, D., Colliot, O., Group, A.S.: Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse. *Medical Image Analysis* **75** (2022) 102219
14. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In Saul, L., Weiss, Y., Bottou, L., eds.: *Advances in Neural Information Processing Systems*. Volume 17., MIT Press (2004)
15. Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.O., Durrleman, S., Colliot, O.: Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies. *Frontiers in Neuroinformatics* **15** (2021) 689675
16. Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C.: The Insight ToolKit image registration framework. *Frontiers in Neuroinformatics* **8** (2014)
17. Thibeau-Sutre, E., Díaz, M., Hassanaly, R., Routier, A., Dormont, D., Colliot, O., Burgos, N.: ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing. *Computer Methods and Programs in Biomedicine* **220** (2022) 106818