Single-Landmark vs. Multi-Landmark Deep Learning Approaches to Brain MRI Landmarking: a Case Study with Healthy Controls and Down Syndrome Individuals

Jordi Malé[1]
jordi.male@salle.url.edu
Yann Heuzé[2]
yann.heuze@u-bordeaux.fr
Juan Fortea[3]
jfortea@santpau.cat
Neus Martínez-Abadías[4]
neusmartinez@ub.edu
Xavier Sevillano[1]
xavier.sevillano@salle.url.edu

[1] HER - Human-Environment Research Group
La Salle - Universitat Ramon Llull
Barcelona, Spain

[2] Univ. Bordeaux, CNRS, Ministère de la Culture
PACEA, UMR 5199, Pessac, France

[3] Sant Pau Memory Unit, Hospital de Sant Pau i la Santa Creu, Barcelona, Spain

[4] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals (BEECA), Facultat de Biologia, Universitat de Barcelona (UB)
Barcelona, Spain

Abstract

Brain dysmorphologies are present in many neurodegenerative, neurodevelopmental and genetic disorders, such as Alzheimer's Disease, schizophrenia, or Down syndrome. Magnetic resonance imaging (MRI) is a widely used tool for diagnosing and monitoring these conditions, but the interpretation of MRI data can be challenging and time-consuming, particularly for large datasets. For this reason, there is a growing interest in developing automatic methods to detect anatomical landmarks in brain MRI data to quantify such dysmorphologies and obtain brain biomarkers to assist in the diagnosis and prognosis of these disorders. In this paper, we propose and evaluate two brain landmarking architectures based on Deep Convolutional Neural Networks (DCNN): an ensemble of single-landmark models, and a multi-landmark model. Both approaches are compared on MRI scans of healthy and Down syndrome subjects. The proposed pipeline comprises several steps: *i)* the preprocessing of the MRI data, involving registration to a common anatomical space, *ii)* the automatic extraction of the Mid-Sagittal Plane (MSP) of the brain, based on a multi-scale search algorithm, and *iii)* the training and evaluation of the DCNN models to detect 8 anatomical landmarks on the MSP. Our results indicate that *i)* the ensemble of single-landmark models is more accurate, achieving average landmarking errors lower than 2mm in healthy subjects, and *ii)* landmarking error is higher in Down syndrome individuals, which suggests that brain dysmorphologies associated with certain disorders require training specific models for accurate landmarking.

1 Introduction

The analysis of medical images plays a fundamental role in the correct diagnosis and treatment of many disorders. Among medical imaging modalities, magnetic resonance imaging (MRI) stands out as a non-invasive, high-resolution, and radiation-free technique that can be used to image a wide range of body tissues and organs. In particular, MRI of the brain is used for the diagnosis and prognosis of multiple genetic and developmental disorders, such as schizophrenia or Alzheimer's disease, among others.

However, the analysis of brain MRI images is a complex and time-consuming task that is usually tackled by expert radiologists and neurologists. This situation has led to the proposal of multiple solutions to perform this interpretation automatically. Key recent contributions in this area include medical image reconstruction, enhancement, segmentation, registration, and landmark detection [15] [2].

Landmark detection plays a crucial role in a wide variety of medical image analysis tasks, as landmark positions are used for the segmentation [7] [5] and registration of different medical images [11][9], together with automatic diagnosis [10]. In the specific area of our study, there has been extensive research on brain MRI segmentation [2] and registration [6]. However, it has been established that additional investigation is necessary to delve deeper into these topics.

However, landmark annotation in brain MRI and, in general, all medical images is often performed manually, which is time-consuming and labour-intensive. For this reason, many computer-assisted landmark detection methods have been developed in recent years and can be classified into two types: traditional and deep learning based.

On the one hand, traditional methods focus on the extraction of invariant features and image filtering [16]. For instance, Asaei *et al.* [1] presented a supervised model that predicts brain landmarks based on the similarity between feature vectors, and Zhang textitet al. [14] proposed a brain landmark detection algorithm based on a random forest regression.

On the other hand, deep learning based methods have obtained more promising results. Yang textitet al. [13] proposed a multi-task learning architecture that automatically detects the anterior and posterior commissures of the brain (AC and PC, respectively). Zhu textitet al. [16] developed a universal anatomical landmark detection model to perform multiple landmark detection tasks with end-to-end training based on a mixed dataset. Edwards textitet al. [3] presented a deep learning model to automatically locate multiple landmarks given 3D brain images.

This paper introduces an end-to-end pipeline -Figure 1- that, using large datasets of brain MRI scans from both healthy individuals and patients, *i)* identifies statistically significant disparities in brain anatomy between distinct groups, *ii)* establishes brain biomarkers with diagnostic potential by leveraging the identified anatomical differences, *iii)* defines a collection of brain landmarks based on the identified biomarkers, *iv)* detects these defined landmarks within brain MRI scans of unseen subjects, *v)* utilizes the detected biomarkers to diagnose new subjects.

In the context of the proposed pipeline, this paper is focused only on the automatic landmark detection process. Specifically, the goal of this paper is to develop accurate two-dimensional brain landmarking algorithms based on DCNN following two opposite approaches: single-landmark, or SL for short (i.e. an ensemble of N DCNNs to detect one landmark each) and multi-landmark (or ML, a single DCNN to detect N landmarks simultaneously). The algorithm was implemented in a 2D context due to the constraints of available annotated data, which exclusively provided landmarks along the Mid-Sagittal Plane (MSP)

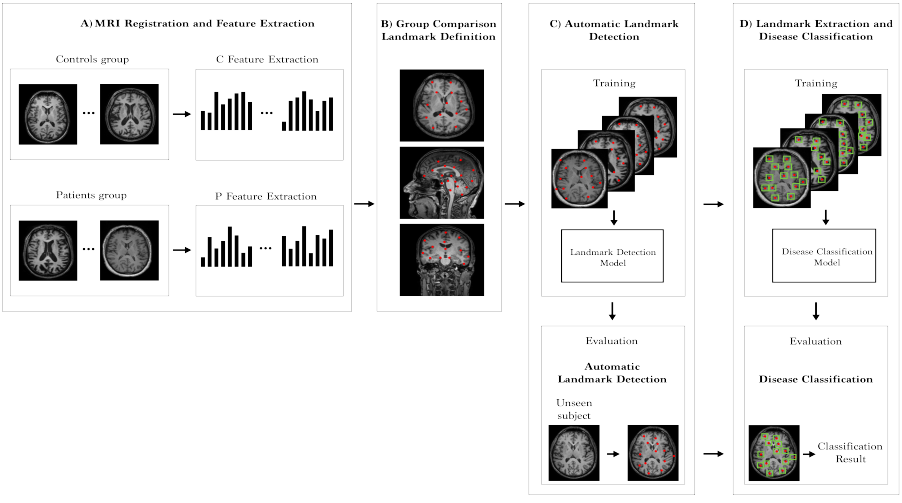


Figure 1: Proposed disease identification framework, based on brain landmarks. It includes (A) MRI Registration and feature extraction, (B) voxel-based statistical group comparison and data-driven landmark definition, (C) automatic landmark detection, (D) landmark information extraction, and landmark-based disease classification.

of the brain. Moreover, we also analysed landmarking accuracy in the brains of healthy controls and individuals with Down syndrome, so as to evaluate the adaptability of the trained models to different brain morphologies.

Our findings revealed detectable differences in prediction errors between the proposed models, with a mean difference of 1 mm. In particular, the proportion of highly accurate landmark detections, with an error of less than 2 mm, doubled when comparing the ensemble of SL models with the ML model. In addition, analysis of the prediction errors on healthy and Down syndrome subjects revealed the need for training specific models to achieve accurate landmarking on patients with particular brain dysmorphologies. Specifically, the ensemble of SL models had a mean landmarking error of 1.84 mm for healthy subjects, while the error increased to 2.25 mm for Down syndrome individuals.

2 Automatic Landmark Detection

Our solution for automatic landmark detection involves three main steps: i) MRI registration and preprocessing, ii) automatic extraction of the MSP, and iii) training and evaluation of Deep Convolutional Neural Network (DCNN) models to detect N anatomical landmarks. An overview of the proposed pipeline is shown in Figure 2.

In the first step, the input volume is in the NIfTI Data Format by Neuroimaging Informatics Technology Initiative [8], and all images are reoriented to Posterior Inferior Left (PIL) orientation. The 2D image slices fed to the DCNN are then subject to further processing, including cropping, histogram matching, and mean-variance normalisation.

The second step involves automatic extraction of the MSP, which is the MRI slice that separates the brain into two almost-identical hemispheres. The MSP is found using a multi-scale search algorithm based on previous work [12], which proposes multiple planes and

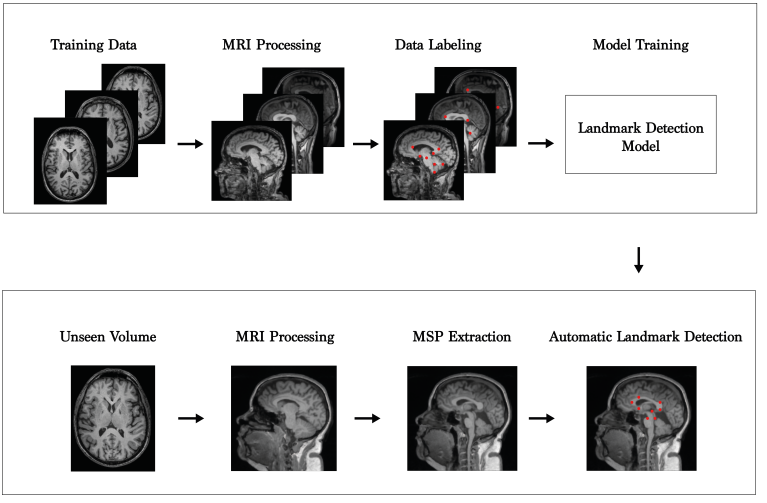


Figure 2: Proposed landmark detection framework, including (1) training of the model using labelled MRIs and (2) testing of the neural network.

evaluates the symmetry defined by the 3D image and the plane itself. Specifically, the objective was to compare our image A with the image B obtained when A is flipped about the plane under evaluation. The method used to quantify the symmetry between these two images is the cross-correlation between images A and B. The equation is as follows:

$$S(A,B) = \frac{(A-\overline{A}\cdot 1)\cdot(B-\overline{B}\cdot 1)}{\|(A-\overline{A}\cdot 1)\|\|(B-\overline{B}\cdot 1)\|} \tag{1}$$

This algorithm focuses on a two-scale process that, first, evaluates the cross-correlation of planes using $\frac{1}{4}$ scale of the volume, and then refines the search using the full volume. At the end of the multi-scale search, the MSP is identified as the plane that maximises brain symmetry on its both sides. The N landmarks will be detected on the MSP.

The third step involves training and evaluating multiple DCNN models for the detection of the N landmarks on the extracted MSP. In this work, the DCNNs models are composed of four blocks comprising each a convolutional layer, a batch normalization layer, a rectified linear unit (ReLU) activation function, and a max-pooling layer, followed by three fully connected layers, together with ReLU activation functions and dropout layers to avoid overfitting. The details and implementation of these models are publicly available for further use (see section 3.2).

In this work, we propose and compare two DCNN-based architectures for detecting N landmarks: a multi-landmark model trained to detect the N landmarks, and an ensemble of N single-landmark models.

On the one hand, single-landmark models are trained with backpropagation using a loss function (L_{SL}) that measures the Euclidean distance between the (x,y)-coordinates of the predicted and the ground truth landmarks (see Equation 2).

$$L_{SL} = \sqrt{(x_{pred}-x_{gt})^2+(y_{pred}-y_{gt})^2} \tag{2}$$

where (x_{pred},y_{pred}) and (x_{gt},y_{gt}) are the coordinates of the predicted and ground truth

landmarks, respectively. Consequently, the output of a SL model is a 2-dimensional vector of the (x,y)-coordinates of the landmark.

On the other hand, the multi-landmark models are trained to minimise a loss function L_{ML} equal to the average of the N Euclidean distances between the coordinates of the predicted and ground truth landmarks (see Equation 3).

$$L_{ML} = \frac{1}{N}\sum_{i}^{N}\sqrt{(x^i_{pred}-x^i_{gt})^2+(y^i_{pred}-y^i_{gt})^2} \tag{3}$$

where (x^i_{pred}, y^i_{pred}) and (x^i_{gt}, y^i_{gt}) are the predicted and ground truth coordinates of the ith of the N landmarks. Thus, the output of a ML model is a $2N$-dimensional vector with the (x,y)-coordinates of the N landmarks.

Consequently, each of the SL models in the ensemble focuses solely on the individual landmark distance, whereas the ML model also considers the inter-landmark distance relationship.

3 Experiments and results

3.1 Data

The SL and ML DCNN models were trained with 1,837 brain MRI scans obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) database (adni.loni.usc.edu)[1]. To assess the performance of the implemented models, we used a completely different dataset of MRI images. In particular, we tested the models with a dataset of 141 brain structural T1 MRI scans obtained with a Philips 3 Tesla X Series Achieva scanner and provided by Hospital Sant Pau Memory Unit (Barcelona, Spain). Interestingly, this test set comprises scans from 87 adult healthy subjects and 54 adults with Down syndrome. The study was approved by the Sant Pau Research Ethics Committees, following the standards for medical research in humans recommended by the Declaration of Helsinki. All participants or their legally authorised representative gave written informed consent before enrolment.

Given the limited availability of manually labelled brain MRI scans, we used the solution proposed in [1] for labelling the training and test set scans, placing $N = 8$ landmarks on the MSP, and serving as the ground truth for our experiments. Figure 3 exhibits the defined landmarks annotated on the mid-sagittal plane of a brain MRI. After labelling [1], all scans were thoroughly reviewed to remove outliers.

3.2 Implementation details

The deep learning models in this study were implemented using PyTorch. The models were initialised using a Gaussian distribution with a mean of 0 and a standard deviation of 0.1. The learning rate was initially set to 0.005 and gradually reduced to a final value of 0.0005 using a gamma factor of 0.95. The scaling operation was performed multiple times, as necessary, until reaching the desired final learning rate at the end of training. This approach ensured a gradual and controlled adjustment of the learning rate throughout the training process,

[1]The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

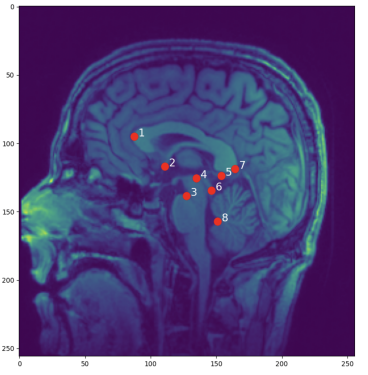


Figure 3: Set of $N = 8$ landmarks located on the MSP.

allowing for better optimisation and convergence of the model. The SL models were trained for 5,000 epochs, and the ML model was trained for 10,000 epochs.

All experiments were conducted on a PC with Intel Core™ i9-10980XE CPU @ 3.00GHz × 36 cores and NVIDIA GeForce RTX 2080 Ti GPU. The code is publicly available on GitHub, https://github.com/jrdimale/DCNN-BrainLandmarking/.

3.3 Evaluation metrics

The evaluation metrics employed for assessing the landmarking accuracy of the SL and ML methods were based on comparing the coordinates of the ground truth and the predicted landmark coordinates in terms of *i)* Euclidean distance (in mm), and *ii)* Successful Detection Rate (SDR), defined as the percentage of landmarks predictions to be within a certain radius of their real position (according to the ground truth), evaluating the accuracy at each of a series of increasing distances (from 1 to 5 mm). Thus, a higher SDR indicates a higher success rate in accurately detecting the landmarks.

3.4 Results

The 141 MRI scans of the test set were subject to automatic landmarking by the ensemble of 8 SL models and the ML model. We measured the accuracy of the landmarking for the healthy subjects (HS) and the Down syndrome individuals (DS) separately. The results are summarised in Table 1, expressed in terms of the Euclidean distance (mean value ± standard deviation) and the SDR metric defined earlier.

The ensemble of SL models performed better than the ML model, achieving an average landmarking error of 1.99±1.94 mm, in comparison to a 3.03±2.13 mm error for the latter. In terms of SDR, the ensemble of SL models achieves 83.86% with a 3mm threshold, compared to 58.86% for the ML model.

Moreover, Table 1 reveals that the landmarking accuracy of both models is higher on the HS than on the DS sample. Taking into account that all our models were trained on the ADNI-1 dataset (which contains a large share of healthy subjects and individuals with mild cognitive impairment), we can infer that accurate landmarking of brains with different dysmorphologies (like those in DS) would require training specific models.

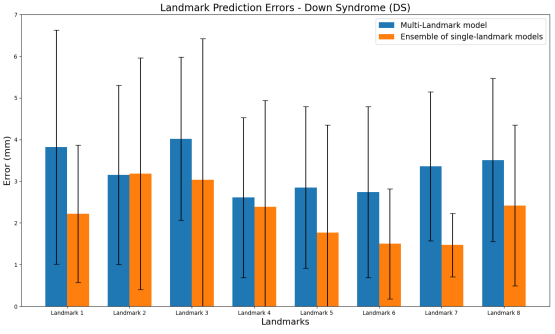


Figure 4: Landmarking error (in mm) on the Down syndrome sample using the ensemble of SL models (orange bars) and the ML model (blue bars), represented as the average value and standard deviation per landmark.

Figure 4 and Figure 5 provide a visual representation of the precision of each individual landmark prediction for the DS and the HS samples, respectively. In both cases, we can see that the ensemble of SL models achieves a lower landmarking error than the ML model in every single landmark (with the possible exception of landmark #2 in the DS sample).

Finally, Figure 6 provides a visual representation of the results of automatic landmarking on three MRI scans of the test set (one HS and two DS individuals). These results reinforce that the ensemble of SL models performs more accurately than the ML model and that this accuracy is higher in the landmarking of HS.

Model	Group	Euclidean distance (mm)	SDR (%)				
			1mm	2mm	3mm	4mm	5mm
ML	DS	3.26 ± 2.15	10.64	31.25	53.47	71.29	84.95
	HS	2.89 ± 2.11	13.36	39.79	62.22	78.59	88.36
SL	DS	2.25 ± 2.35	25.92	62.73	80.32	87.96	90.97
	HS	1.84 ± 1.63	31.75	68.53	86.06	92.09	96.12

Table 1: Landmarking error in terms of Euclidean distance (mean value $\pm$ standard deviation) and SDR with 1mm to 5mm error thresholds for both the SL and ML models and sample groups (HS and DS).

4 Conclusions and further work

In this work, we investigated the accuracy of deep learning models for the automatic detection of landmarks on a given 2D MRI sagittal slice of the brain. Two approaches have been compared: an ensemble of single-landmark models, and a multi-landmark model. Moreover, the accuracy of both approaches has been evaluated on MRI scans of subjects with different brain morphologies (healthy subjects and individuals with Down syndrome).

The paper's main objective is to achieve precise landmarking and to begin investigating the distinction between healthy subjects and those with Down Syndrome or other illnesses.

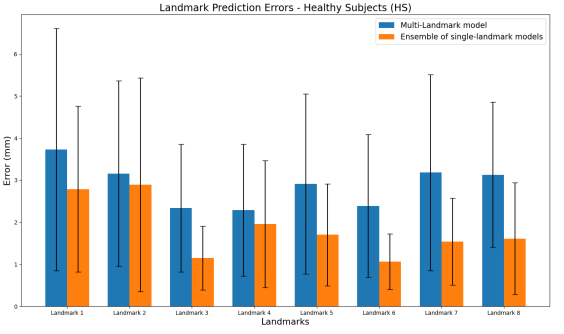


Figure 5: Landmarking error (in mm) on the healthy subjects sample using the ensemble of SL models (orange bars) and the ML model (blue bars), represented as the average value and standard deviation per landmark.

Nonetheless, directly comparing our method to state-of-the-art techniques might not be entirely equitable, as these methods utilize uniform datasets and do not specifically focus on identifying potential diagnostic biomarkers.

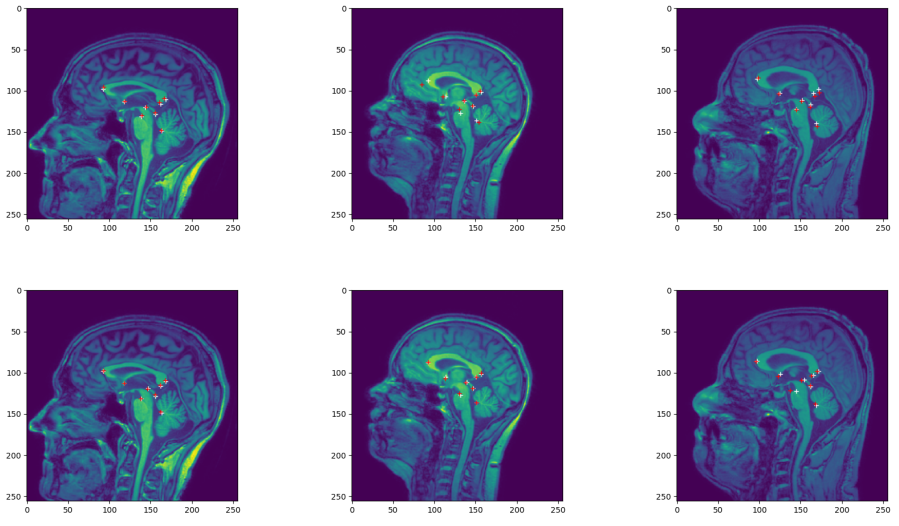
As our experimental results reveal, the ensemble of single-landmark models performed better for all landmarks, obtaining more accurate results than the multi-landmark model. This suggests that it is preferable to have a model that focuses on only the loss function of the Euclidean distance of a single landmark, rather than focusing on the average loss functions of N-landmarks and the relative position between them, as the ML model does.

Furthermore, we have also analysed the performance of both landmarking models on test data from different groups. This analysis aims at evaluating the ability to generalize landmark predictions to unseen brain morphologies. The fact that both models perform better on the healthy subjects sample than on the Down syndrome sample highlights the need for new models to be trained specifically to predict landmarks for groups of patients with particular brain morphologies.

Future research directions include several aspects to improve the implemented models and extend their applicability to different patient groups. First, the exploration of new datasets would contribute to improving the performance and generalization of the models. In addition, the incorporation of data augmentation techniques may improve the robustness of the models and their ability to handle variations in the input data. On the other hand, the differences observed between groups highlight the need for future research focused on defining new biomarkers based on regions of the brain that are linked to specific diseases, rather than relying on pre-defined landmarks based on prior knowledge. Moreover, our work aims to develop a three-dimensional model to predict these newly defined landmarks and, ultimately, to train a classifier to diagnose based on the features of the extracted biomarkers.

Acknowledgements

This work was partly supported by Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) of the Generalitat de Catalunya (2021 SGR01396, 2021 SGR00706), the Spanish Ministry of Science, Innovation, and Universities under grant PID2020-113609RB-C21



(a) (b) (c)

Figure 6: Results of automatic landmarking on three brain MRI scans of the test set. The ground truth landmarks are depicted as white crosses, and the automatically detected landmarks as red crosses. Images on the top row correspond to landmarking by the ML model, and images on the bottom row by the ensemble of SL models. Column (a) corresponds to a healthy subject, and columns (b) and (c) correspond to Down syndrome individuals.

/AEI/10.13039/501100011033, the Fondation Jerome Lejeune under grant 2020b cycle-Project No.2001, and the Joan Oró grant (FI 2023) from the DRU of the Generalitat de Catalunya and the European Social Fund (2023 FI-1 00014). Data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[2].

References

[1] Ali Asaei. *Brain MRI landmark identification and detection*. PhD thesis, 2015. URL http://hdl.handle.net/20.500.12648/650.

[2] A. L. W. Bokde, S. J. Teipel, Y. Zebuhr, G. Leinsinger, L. Gootjes, R. Schwarz, K. Buerger, P. Scheltens, H.-J. Moeller, and H. Hampel. A new rapid landmark-based regional MRI segmentation method of the brain. *Journal of the Neurological Sciences*, 194(1):35–40, 2002. ISSN 0022-510X. doi: https://doi.org/10.1016/S0022-510X(01)00667-0. URL https://www.sciencedirect.com/science/article/pii/S0022510X01006670.

[3] Christine A. Edwards, Abhinav Goyal, Aaron E. Rusheen, Abbas Z. Kouzani, and Kendall H. Lee. Deepnavnet: Automated landmark localization for neuronavigation. *Frontiers in Neuroscience*, 15, 2021. ISSN 1662-453X. doi: 10.3389/

[2]The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report (see http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

fnins.2021.670287. URL https://www.frontiersin.org/articles/10.3389/fnins.2021.670287.

[4] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: a review. *Physics in Medicine Biology*, 65(20):20TR01, oct 2020. doi: 10.1088/1361-6560/ab843e. URL https://dx.doi.org/10.1088/1361-6560/ab843e.

[5] Nicolás Gaggion, Lucas Mansilla, Diego H Milone, and Enzo Ferrante. Hybrid graph convolutional neural networks for landmark-based anatomical segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 600–610. Springer, 2021.

[6] Dong Han, Yaozong Gao, Guorong Wu, Pew-Thian Yap, and Dinggang Shen. Robust anatomical landmark detection for mr brain image registration. volume 17, pages 186–93, 09 2014. ISBN 978-3-319-10403-4. doi: 10.1007/978-3-319-10404-1_24.

[7] Bulat Ibragimov, Robert Korez, Bostjan Likar, Franjo Pernus, Lei Xing, and Tomaž Vrtovec. Segmentation of pathological structures by landmark-assisted deformable models. *IEEE Transactions on Medical Imaging*, PP:1–1, 02 2017. doi: 10.1109/TMI.2017.2667578.

[8] Neuroimaging Informatics Technology Initiative. Nifti data format, 2020. URL https://nifti.nimh.nih.gov.

[9] Thomas Lange, Nils Papenberg, Stefan Heldmann, J. Modersitzki, Bernd Fischer, Hans Lamecker, and Peter Schlag. 3d ultrasound-ct registration of the liver using combined landmark-intensity information. *International journal of computer assisted radiology and surgery*, 4:79–88, 01 2009. doi: 10.1007/s11548-008-0270-1.

[10] Mingxia Liu, Jun Zhang, and Ehsan Adeli. Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical Image Analysis*, 43, 10 2017. doi: 10.1016/j.media.2017.10.005.

[11] R. Revathy, S. Kumar, V. Reddy, and Bhavana V. *Medical Image Registration Using Landmark Registration Technique and Fusion*, pages 402–412. 01 2020. ISBN 978-3-030-37217-0. doi: 10.1007/978-3-030-37218-7_46.

[12] Guilherme Ruppert, Leonid Teverovskiy, Chen-Ping Yu, Alexandre Falcão, and Yanxi Liu. A new symmetry-based method for mid-sagittal plane extraction in neuroimages. pages 285–288, 03 2011. doi: 10.1109/ISBI.2011.5872407.

[13] Xulei Yang, Wai Teng Tang, Gabriel Tjio, Si Yong Yeo, and Yi Su. Automatic detection of anatomical landmarks in brain MR scanning using multi-task deep neural networks. *Neurocomputing*, 396:514–521, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.10.105. URL https://www.sciencedirect.com/science/article/pii/S0925231219304874.

[14] Jun Zhang, Yue Gao, Yaozong Gao, and Brent Munsell. Detecting anatomical landmarks for fast alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging*, 35:1–1, 06 2016. doi: 10.1109/TMI.2016.2582386.

[15] S. Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S. Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L. Prince, Daniel Rueckert, and Ronald M. Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820–838, may 2021. doi: 10.1109/jproc.2021.3054390. URL https://doi.org/10.1109%2Fjproc.2021.3054390.

[16] Heqin Zhu, Qingsong Yao, Li Xiao, and S. Kevin Zhou. You only learn once: Universal anatomical landmark detection. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 85–95. Springer International Publishing, 2021. doi: 10.1007/978-3-030-87240-3_9. URL https://doi.org/10.1007%2F978-3-030-87240-3_9.