Machine Learning with Applications 22 (2025) 100790

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Deep learning-based 3D reconstruction of dentate nuclei in Friedreich's ataxia from T2*weighted MR images

Trushal Sardhara [a], Ravi Dadsena [a,b], Roland C. Aydin [c,d], Ralf-Dieter Hilgers [e], Leon Horn [a], Jörg B. Schulz [a,b], Kathrin Reetz [a,b,*], Sandro Romanzetti [a,b,$], Imis Dogan [a,b,$], the FACROSS study group

[a] *Department of Neurology, RWTH Aachen University, Aachen, Germany*
[b] *JARA-BRAIN Institute Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany*
[c] *Institute for Continuum and Material Mechanics, Hamburg University of Technology, Hamburg, Germany*
[d] *Institute of Material Systems Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany*
[e] *Department of Medical Statistics, RWTH Aachen University, Aachen, Germany*

ARTICLE INFO

Keywords:
Automated reconstruction
Neuroimaging markers
Neurodegeneration
Longitudinal analysis
Disease progression

ABSTRACT

Dentate nucleus (DN) degeneration is a key neuropathological feature in Friedreich's ataxia (FRDA), and its accurate quantification is critical for understanding disease progression. However, its visualization and volumetry require iron-sensitive imaging techniques and time-consuming segmentation procedures, posing challenges for conventional ML approaches due to small datasets typical of rare diseases. We present a transfer learning–based machine learning pipeline for automated DN segmentation that directly uses standard T2*-weighted Magnetic Resonance Imaging (MRI), which highlights the DN without additional processing, and is designed to perform robustly with limited annotated data. Using 38 manually labeled subjects (18 FRDA, 20 controls), the model was validated via five-fold cross-validation and an independent hold-out test set, achieving Dice scores of 0.81–0.87 and outperforming classical atlas-based methods. Pretraining improved performance by ∼10% in patients and >5% in controls. Applied to 181 longitudinal scans from 33 FRDA patients and 33 controls, the model revealed significantly reduced DN volumes in FRDA, with reductions correlating with disease duration and clinical severity over time. Our approach provides a scalable and reproducible segmentation framework, requiring minimal annotated data and no preprocessing, while demonstrating robust performance across cross-validation and independent testing. Additionally, it enables the first longitudinal volumetric analysis of DN in FRDA using standard T2*-weighted MRI, demonstrating its practical utility for monitoring neurodegenerative changes. Overall, this work illustrates how transfer learning can overcome data scarcity in rare diseases and provides a robust methodology for automated MRI segmentation in both research and clinical applications.

1. Introduction

Friedreich's ataxia (FRDA) is a rare neurodegenerative disorder caused by mutations in the *FXN* gene that lead to reduced levels of the mitochondrial protein frataxin (Campuzano et al. 1996). This deficiency results in progressive loss of coordination, muscle weakness, and other non-neurological complications, including cardiomyopathy. As the disease progresses, it causes substantial disability, often necessitating mobility aids in young adulthood. Degeneration of the dentate nuclei (DN) plays a key role in the development of the primary clinical symptom ataxia in patients with FRDA, alongside degeneration of the spinal cord, peripheral nerves, and other specific brain regions (Dogan et al., 2019; Koeppen et al., 2007).

The DN, located within the cerebellum, are iron-rich regions and

---

* Corresponding author: Prof. Dr. Kathrin Reetz, Department of Neurology, RWTH Aachen University, Pauwelsstrasse 30, 52074 Aachen, Germany; Phone: +49-241-80 89600

E-mail addresses: tsardhara@ukaachen.de (T. Sardhara), rdadsena@ukaachen.de (R. Dadsena), Roland.aydin@tuhh.de (R.C. Aydin), rhilgers@ukaachen.de (R.-D. Hilgers), Leon.horn@rwth-aachen.de (L. Horn), jschulz@ukaachen.de (J.B. Schulz), kreetz@ukaachen.de (K. Reetz), sromanzetti@ukaachen.de (S. Romanzetti), idogan@ukaachen.de (I. Dogan).

[$] Senior principal investigator, equal contribution

https://doi.org/10.1016/j.mlwa.2025.100790
Received 12 September 2025; Received in revised form 24 October 2025; Accepted 5 November 2025

particularly susceptible to FRDA (Koeppen et al., 2007). Histopathological studies have demonstrated significant atrophy of large glutamatergic neurons, gliosis, and iron redistribution within the DN (Koeppen & Mazurkiewicz, 2013). Consequently, volumetric analysis of DN is crucial to understand the progression of FRDA. In vivo imaging of DN, however, is not common in routine magnetic resonance imaging (MRI) and requires sequences that are sensitive to tissue magnetic susceptibility due to its high iron content. T2*-weighted imaging exploits iron-induced local field inhomogeneities and provides a better contrast for delineating the DN, allowing visualization of its boundaries and microstructural changes that are obscured on T1- or T2-weighted MRI (Fig. 1). Beyond conventional T2*-weighted imaging, other advanced iron sensitive techniques, such as susceptibility-weighted imaging (SWI) and quantitative susceptibility mapping (QSM), can further enhance DN assessment and have been used in different neurodegenerative diseases, including FRDA (Deistung et al., 2016; Ravanfar et al., 2021). SWI provides valuable information on the iron distribution within tissues by combining magnitude and phase information of T2*-weighted images (Haacke et al., 2004). QSM additionally allows improved anatomical delineation and in vivo quantification of iron concentration. Despite its advantages, QSM requires extensive post-processing steps to obtain the final susceptibility maps, which can be used to perform segmentations (Liu et al., 2015). T2*-weighted imaging, on the other hand, is an easily accessible technique and widely implemented in clinical and research protocols. It enables sufficient and effective detection of pathological changes in iron-rich regions without requiring advanced processing steps as involved in SWI and QSM. Importantly, T2*-weighted imaging is a well-validated and widely used technique in clinical practice that can be fairly easily optimized for acquisition-related variations and magnetic field inhomogeneities (Chavhan et al., 2009).

Manual segmentation of DN remains the gold standard for accurate delineation. However, this approach is highly time-intensive and susceptible to inter-observer variability and biases, limiting its scalability and reproducibility. Atlas-based methods have been widely used as an efficient alternative (Diedrichsen et al., 2009, 2011; He et al., 2017). However, single-atlas approaches are unable to capture the considerable inter-subject variability in brain anatomy, often leading to reduced accuracy (Doan et al., 2010).

In recent years, machine learning (ML)–based methods have been increasingly applied to medical image segmentation, providing automated alternatives to labor-intensive manual procedures and often outperforming traditional atlas-based approaches (Gibbons et al., 2023). Foundational ML models for medical image segmentation have been developed to handle a wide range of modalities and tasks (B. Azad et al., 2023; Ma et al., 2024), with convolutional neural networks demonstrating particular promise due to their ability to capture complex spatial patterns. In this context, the U-Net architecture has emerged as a highly effective solution for medical image segmentation (Ronneberger et al., 2015). Given the inherently three-dimensional nature of brain MRI data, leveraging a 3D U-Net (Çiçek et al., 2016) enhances segmentation accuracy by capturing spatial relationships across volumetric datasets, improving both efficiency and reliability for DN segmentation. Nevertheless, ML-based segmentation methods face challenges (Bottou, 2014; Burkart & Huber, 2021; Kotsiantis et al., 2006), among which the most significant ones are the scarcity of ground-truth values and insufficient training data (Brodley et al., 2012). This challenge is especially critical in rare diseases such as FRDA, where manual annotation is prohibitively time-consuming and data availability is low. Transfer learning has been proposed as a powerful strategy to mitigate data scarcity by pretraining models on large datasets from related domains and fine-tuning them on smaller, disease-specific datasets. This technique has demonstrated significant potential in different scientific and engineering applications including medical imaging (Kim et al., 2022; Shafighfard et al., 2025). For example, Alshardan et al. successfully applied transfer learning for semantic segmentation of brain tumor tissues using MRI, demonstrating its effectiveness in studies with limited sample sizes (Alshardan et al., 2024).

Despite these advances, the application of ML techniques to the DN remains limited, and studies in the context of FRDA are particularly scarce. Recent approaches include Beliveau et al. (Beliveau et al., 2021), who utilized SWI, while Chai et al. (Chai et al., 2022) and Shiraishi et al. (Shiraishi et al., 2025) implemented QSM to segment iron-rich DN automatically. Bermudez Noguera et al. (Bermudez Noguera et al., 2019) further explored ML-based DN segmentation across multiple MRI modalities, including T1-, T2-, and diffusion-weighted imaging, highlighting the potential of multimodal approaches. However, most existing studies rely on specialized imaging or large datasets, limiting their applicability to rare diseases with scarce annotated data.

In this study, we make two key contributions. First, we present a transfer learning–based segmentation framework that enables accurate DN delineation using a small number of commonly acquired T2*-weighted MRI scans, a clinically accessible modality that provides inherent DN contrast without requiring additional preprocessing. This strategy mitigates the need for large annotated datasets while maintaining robust segmentation performance. Second, we apply this framework to a clinical cohort of FRDA patients and controls, providing both cross-sectional and longitudinal analyses of DN volumes over up to four years of follow-up. To our knowledge, this is the first study to demonstrate longitudinal DN volumetry in FRDA using T2*-weighted MRI, thereby providing a robust and scalable method for automated DN analysis in rare neurodegenerative disorders.

2. Methods

2.1. MRI data acquisition

33 individuals with genetically confirmed FRDA and 33 age- and sex-matched healthy controls were included in this study. Patients were enrolled as part of their annual visits for the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) (Dogan et al., 2019; Reetz et al., 2015, 2016, 2021) at the University Hospital RWTH Aachen. They were asked to participate in an MRI sub-study, which was

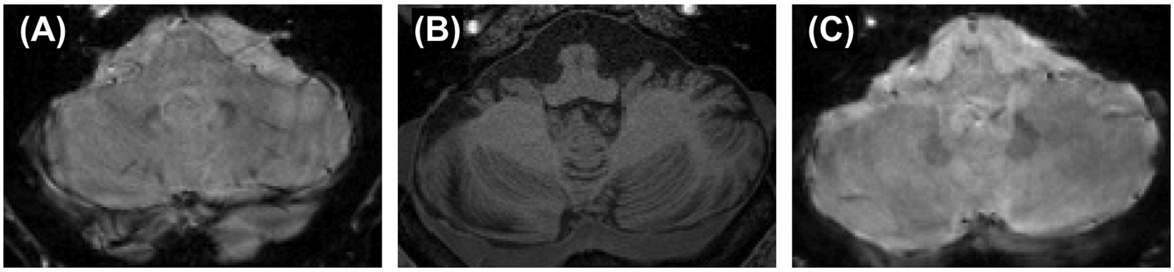

Fig. 1. Processed MR image slices of a Friedreich's ataxia patient in different modalities: **(A)** Raw T2*-weighted MRI slice, **(B)** processed T1-weighted MRI slice, and **(C)** processed T2*-weighted MRI slice. These images highlight the structural differences captured by each modality, which are relevant for dentate nuclei segmentation.

approved by the local ethics committee (EK057/10, EK083/15). Longitudinal data were available for 25 patients and 19 controls, each with at least one follow-up visit, yielding a total of 181 MRI datasets over four years (Table 1).

MRI data were acquired using a 3T PRISMA whole-body MR scanner (Siemens Healthineers, Erlangen, Germany) with a 64-channel head–-neck receiver coil. T2*-weighted images were obtained using a 3D multi-echo FLASH sequence with the following parameters: TR = 30 ms, 11 echoes ranging from TE = 2.16 to TE = 23.7 ms, resolution 1 × 1 × 2 mm^3. Only images from the last echo were employed for manual segmentations to enhance the detection of the dentate nuclei. High-resolution T1-weighted MR images were acquired using a magnetization-prepared rapid gradient echo sequence with the following parameters: TR = 2400 ms, TE = 2.36 ms, TI = 1000 ms, flip angle 8 degrees, 208 sagittal slices, resolution 0.8 × 0.8 × 0.8 mm^3.

2.2. *Data preprocessing and manual annotation*

Due to the different voxel resolutions of T2*-weighted and T1-weighted images, we first resampled and registered both image modalities to achieve the same isotropic resolution of 0.8 mm in all directions. Fig. 1 shows one slice from the original T2*-weighted image and the processed T1-weighted and T2*-weighted images.

T1-weighted images were utilized for two main objectives: isolating the cerebellum region to obtain a final volume of interest and implementing a pretraining task. To obtain the volume of interest, i.e., the cerebellum, we performed whole-brain segmentation for all 181 MRI datasets using the fully automated Fastsurfer (v 2.3.3) (Henschel et al., 2020) deep learning-based neuroimaging pipeline, which utilizes only T1-weighted images. Once the binarized cerebellum volume masks were extracted, we cropped smaller volumes of size 192 × 192 × 192 voxels from the original MRI T2*-weighted image data. Supplementary Figure S1 illustrates the full workflow of the data preprocessing step. These cropped T2*-weighted volumes were used throughout the study to train the ML models. For pretraining, we selected cerebellum white-matter (WM) classes (class labels 7 and 46 for left and right cerebellar WM, respectively) from the whole-brain segmentation. Additionally, the estimated total intracranial volume (eTIV) was derived from Fastsurfer to account for individual variability in brain size in subsequent statistical analyses.

After these preprocessing steps, we normalized T2*-weighted images to T1-weighted images and manually annotated the DN using T2*-weighted images acquired at the longest echo time. The extended echo time in T2*-weighted images significantly enhanced image contrast in the DN region due to their elevated iron concentration. The labeling procedure was performed blinded to participant group and carried out using ITK-SNAP (v 3.6.0) in all three planes (axial, sagittal, coronal) by careful tracing and scrutiny of the signal intensity in and along the DN surface region. This rigorous classification procedure ultimately ensured estimation of DN volumes in each hemisphere as precisely as possible and as a complete macroscopic structure visible on MRI, i.e., the cortical gray matter ribbon together with the enclosed white matter. Due to the limited resolution of MRI, it is not possible to reliably isolate the thin gray matter ribbon of the DN that is typically defined in microscopic histological descriptions. Supplementary Figure S2 illustrates the manual annotation of DN volumes on T2*-weighted data.

Table 1
Characteristics of the study groups at baseline and data availability.

		FRDA patients (n = 33)	**Controls** (n = 33)	p
Sex (female/male)		17/16	21/12	0.319
Age (years)		34.2 ± 11.8	32.2 ± 11.8	0.488
Age of onset (years)		16.3 ± 7.0	-	
Disease duration (years)		18.3 ± 9.0	-	
GAA repeats*	allele 1	520.3 ± 192.8	-	
	allele 2	842.1 ± 171.5	-	
SARA total score		19.2 ± 8.7	-	
ADL total score		13.3 ± 6.5	-	
Follow-up scans, time (years)	1 year	23; 1.02 ± 0.14	17; 1.08 ± 0.15	0.176
	2 years	20; 1.96 ± 0.18	12; 2.23 ± 0.23	<0.001
	3 years	14; 3.05 ± 0.15	11; 3.08 ± 0.22	0.658
	4 years	11; 4.02 ± 0.12	7; 4.41 ± 0.14	<0.001
total number		101; 1.48 ± 1.37	80; 1.37 ± 1.47	

Data are presented as mean ± standard deviation; n, number of participants; p, p-value (statistical significance was set at $p < 0.05$); FRDA, Friedreich's ataxia; SARA, Scale for the Assessment and Rating of Ataxia; ADL, Activities of Daily Living. *GAA repeat length missing in one patient, point mutation in another patient.

2.3. *Semantic segmentation*

2.3.1. *Machine learning architecture*

Given that MRI data are volumetric, leveraging spatial information is important. Therefore, we employed a 3D convolutional neural network for our study. Specifically, we implemented a 3D U-Net (Çiçek et al., 2016) and modified it with various kernel sizes, residual connections (Lee et al., 2014), and different numbers of encoding blocks as hyperparameters. Additionally, we incorporated three significant modifications to enhance the network's performance:

1. A self-attention module (Oktay et al., 2018) was added between the encoder and decoder layers to help the network focus on important regions, such as DN.
2. A bi-directional ConvLSTM (R. Azad et al., 2019) layer was included in the skip connections to capture more discriminatory information, thereby improving segmentation precision.
3. A deep supervision strategy (Lee et al., 2014) was implemented to optimize learning at multiple scales.

Supplementary Figure S3-A illustrates the modified U-Net architecture used in this study. To assess the impact of each architectural modification, we conducted an ablation study comparing the baseline network (without modifications) against variants incorporating individual modifications, combinations of these modifications, and the full proposed architecture integrating all modifications.

2.3.2. *Pretraining procedure*

To address the challenge of limited annotated brain volumes, we pretrained our ML model using paired data prepared with Fastsurfer. The training data consisted of all 181 cropped T2*-weighted images, with automatic segmentation calculated using Fastsurfer serving as the "ground truth". This model was optimized to segment WM class labels from the cerebellum (Figure S3-B – Step 1). Theoretically, the model can be optimized for any task before being used for the final downstream task. We chose cerebellar WM as the pretraining label because the DN, although a gray matter structure, is embedded within this region and is captured by the FastSurfer segmentation output. Pretraining on WM allows the network to learn contextual features surrounding the DN, improving subsequent DN segmentation. Once our models performed optimally for the WM segmentation task, we used these model weights with the same architecture to train the model for segmenting DN in the second step (Figure S3, (B) – Step 2).

2.3.3. *Training procedure*

All ML models were implemented using PyTorch (v 2.3.0) framework and trained on one H100 GPU. To achieve the final segmentation, we trained three separate models for sagittal, axial, and coronal views. Unlike simple augmentation (e.g., flipping), each orientation emphasizes distinct structural relationships and provides unique receptive field alignment. By combining their predictions through weighted averaging, we reduced orientation-specific biases and achieved more robust segmentations. To optimize each model, images were cropped into smaller patches (128 × 128 × 128) with a 64-pixel stride using a sliding window technique. Training was conducted on these 3D image patch volumes. Inspired by Milletari et al. (Milletari et al., 2016), we used a squared Dice loss layer with a smoothness value to avoid division by zero in all three ML models.

Hyperparameters, including the initial learning rate, size, optimizer configuration, and loss function, were guided by prior work on U-Net architectures in medical image segmentation (Çiçek et al., 2016; Milletari et al., 2016) and initially tuned on a small pilot subset of the data to identify stable ranges and ensure convergence. The batch size was chosen based on computational capacity. The final settings (see Table S2) were then fixed and applied consistently across all models and folds to prevent overfitting to the main dataset. The learning rate was reduced by a factor of 10 if no improvement in loss was observed for ten consecutive epochs. Once each model was optimized for its respective view, the predictions from each model were combined using a weighted average. We assigned weights of 0.2, 0.6, and 0.2 to the sagittal, coronal, and axial views, respectively. These weights were determined by testing various combinations for maximum performance. Supplementary Table S2 summarizes the detailed training parameters for all models.

2.3.4. *Validation strategy*

From a total of 181 MRI datasets, for the selected 38 datasets (18 FRDA, 20 controls) acquired with the longest echo time, we adopted a two-tier validation approach to ensure both internal robustness and external generalizability. First, we set aside an independent hold-out test set consisting of 4 (2 FRDA, 2 controls) manually annotated MRI scans (≈10 % of the data), which was reserved exclusively for final performance assessment and was not used at any stage of model development, training, or tuning. The remaining 34 scans (16 FRDA, 18 controls) were used for training and internal validation. To evaluate model stability within this cohort, we performed five-fold cross-validation. In each fold, approximately 80 % of the scans were used for training and 20 % for validation, with subjects partitioned to ensure non-overlapping data across folds. This procedure ensured that every subject contributed to validation exactly once, reducing bias from data reuse.

Performance metrics were averaged across folds and reported as mean ± standard deviation in supplementary materials (see Table S3), providing an estimate of model stability and variance across subsets. The independent test set was used exclusively for final performance assessment. This design ensured that the reported test results reflected true generalization to unseen data, while the cross-validation analysis provided additional evidence of robustness and stability despite the small sample size.

2.3.5. *Atlas-based segmentation*

To compare our ML-based segmentation results with the commonly used method, we implemented an atlas-based DN segmentation approach using the probabilistic cerebellar atlas of Diedrichsen et al. (Diedrichsen et al., 2009). The atlas, defined in MNI space with 1 × 1 × 1 mm^3 voxel resolution, was first conformed to ensure consistency in dimensions and resolution. Non-linear registration to each subject's T2*-weighted MRI data was performed using the symmetric normalization algorithm implemented in Advanced Normalization toolbox (antspyx v0.4.2), after which the DN labels were propagated to the individual subject space. While the atlas-based segmentation method may not be the most accurate, it provides a conventional anatomical reference against which the performance of the machine learning models can be compared.

2.4. *Metrics*

In this study, ground truth data (manually annotated) were available for 38 brain volumes, allowing us to calculate absolute error metrics to evaluate the performance of our segmentation models. We used the Dice score (DS) (Milletari et al., 2016), which measures the similarity between two sets of (pixel or voxel) data, in this case, the segmented DN and ground truth DN datasets. Its value ranges from 0 to 1, with 1 indicating perfect overlap between the two datasets. The Dice score accounts for imbalanced datasets, as in our case, where the number of dentate nuclei region voxels is much less than the other class, which is the background. We calculated the Dice score by

$$DS = \frac{2\,TP}{2\,TP + FN + FP}.$$

where TP is the number of true positive voxels, FN is the number of false negative voxels, TN is the number of true negatives, and FP is the number of false positive voxels.

2.5. *Statistical group analysis*

The ML-based segmentation method was applied to all 181 available datasets to derive volumes of segmented DN at baseline and follow-up visits over four years. Cross-sectional group differences were assessed using two-sample *t*-tests and Cohen's d as an effect size measure (Montgomery et al., 2021). Associations of DN volumes with disease duration, age of onset, GAA repeat length, and clinical severity at baseline were tested with Spearman's rho. The Scale for the Assessment and Rating of Ataxia (Schmitz-Hübsch et al., 2006) (SARA; 40-point scale) was used to measure ataxia severity, and Activities of Daily Living (Subramony et al., 2005) (ADL; 36-point scale) for functional impairment, with higher scores indicating greater clinical impairment. Longitudinal changes in DN volumes were analyzed using linear mixed effect modelling (restricted-maximum-likelihood estimation method) with random effects on slope (i.e., time in years) by including baseline DN volumes and time as fixed main effects, and interaction terms between group and time. Associations over time between clinical impairment and DN volumes were tested in separate models using clinical measures as predictors. DN volumes were corrected for eTIV in all analyses (expressed as percentage of the estimated total intracranial volume, %eTIV). Statistical analyses were performed in SPSS (v29.0.1.1), and the threshold for significance was set at $p \leq 0.05$ (two-sided).

3. Results

3.1. *Automatic segmentation using ML*

To evaluate the impact of our transfer learning approach, two segmentation pipelines were trained on manually annotated data: one with pretraining and one without. Without pretraining, the average Dice score for whole DN segmentation was 0.76 ± 0.01 in FRDA patients and 0.81 ± 0.02 in controls. The pretrained model was then used to segment DN volumes across the full dataset, yielding consistently high Dice scores and demonstrating strong agreement between predicted and manual segmentations. Pretraining improved performance, increasing Dice scores to 0.83 ± 0.02 in FRDA patients (10 % increase) and to 0.85 ± 0.01 in controls (5 % increase; Table 2). Slightly higher Dice scores were observed for the left DN compared to the right DN in both FRDA patients and controls. However, these differences were not statistically significant ($p = 0.56$ for patients; $p = 0.13$ for controls) and are more likely attributable to minor annotation or image variability than to a true lateralization effect. Additionally, we calculated segmentations of the test dataset using the base ML-based method without any architectural modifications and atlas-based (Diedrichsen et al., 2009) method and compared the Dice scores to our ML-based segmentation method with proposed modifications in Table 2. Overall, our ML-based method consistently outperformed the base ML-based model and the classical atlas-based segmentation approach. Specifically, architectural

Table 2
Dice scores of DN segmentation using our ML-based method, ML-based method without architectural modifications, and atlas-based method for comparison.

		Whole DN	Left DN	Right DN
ML-based method (with proposed modifications)	FRDA	0.83 ± 0.02	0.84 ± 0.03	0.81 ± 0.01
	Controls	0.85 ± 0.01	0.87 ± 0.01	0.83 ± 0.02
ML-based method (no modifications)	FRDA	0.78 ± 0.03	0.77 ± 0.03	0.78 ± 0.04
	Controls	0.83 ± 0.02	0.84 ± 0.01	0.82 ± 0.02
Atlas-based method	FRDA	0.22 ± 0.12	0.23 ± 0.10	0.21 ± 0.15
	Controls	0.35 ± 0.05	0.38 ± 0.07	0.33 ± 0.08

Dice scores for the DN segmentation reflect the model's accuracy in segmenting each region from T2*-weighted MRI data for both healthy controls and FRDA patients. Higher Dice score is better. ML, Machine learning; FRDA, Friedreich's ataxia; DN, Dentate nucleus.

modifications improved performance, increasing Dice scores by 6.4 % in FRDA patients and 2.4 % in controls (Table 2). A detailed ablation analysis, including results for the baseline model, individual and combined modifications, as well as the fully modified architecture, is provided in Supplementary Table S1. Additionally, ML-based segmentation achieved Dice scores that were 2–3 times higher than those of the atlas-based method, which ranged from 0.20 to 0.38 (Table 2). Representative segmentation results are shown in Fig. 2, highlighting the accurate delineation of the dentate nuclei from T2*-weighted MR images by the ML-based segmentation model.

3.2. *Group analysis and clinical correlations*

To determine volumetric differences in the DN between FRDA patients and controls, DN volumes were segmented from 181 brain MRI datasets using the above-described ML method. Cross-sectionally, DN volumes were significantly smaller in FRDA patients compared to controls at baseline (mean ± SD: 1062 ± 290 mm^3 vs. 1261 ± 275 mm^3 adjusted for eTIV, $p = 0.006$) and each follow-up visit, with medium to large effect sizes observed at each time point (Cohen's $d = 0.704$ to 1.39, Fig. 3A). Longitudinal analysis based on linear mixed effect modelling did not show significant changes in DN volumes over time or group-by-time interactions.

At baseline, we did not find significant associations between DN volumes and measures of disease severity in FRDA patients; however, baseline DN volumes were correlated with changes in ADL scores after four years ($r = 0.668$, $p = 0.025$). Longitudinally, decreases in right DN volume were significantly associated with longer disease duration (estimate [SE]: −1.57 [0.76] mm^3, $p = 0.041$), indicating an average volume loss of about 1.5 mm^3 per year of disease duration (Fig. 3B). Similarly, worsening in SARA scores (–1.97 [0.80] mm^3, $p = 0.016$) and ADL scores (–2.86 [1.07] mm^3, $p = 0.009$) were related to a decrease of approx. 2 to 3 mm^3 in right DN volumes, respectively (Fig. 3B).

4. Discussion

In this study, we developed an ML-based method for segmenting the DN using a very small T2*-weighted MR images dataset and presented the first longitudinal volumetric analysis of up to four years of DN changes in FRDA. Augmented with transfer learning, our approach achieved high segmentation accuracy for DN volumes in patients with FRDA and control participants. It produced state-of-the-art results competitive with advanced ML-based methods, while requiring significantly fewer annotated datasets and utilizing widely available T2*-weighted images. The substantial improvement over the atlas-based baseline segmentation method underscores the clinical and research utility of the proposed ML-based segmentation approach. Furthermore, volumetric analysis of DN revealed significantly reduced DN volumes in FRDA patients compared to controls, with reductions correlating with disease duration and progression of clinical severity over time.

To assess model stability, we performed five-fold cross-validation on the training/validation cohort. Cross-validation results closely matched independent test set performance, demonstrating strong generalization and robustness even with limited data. The ablation study confirmed that the proposed architectural modifications contributed to an average Dice score improvement of ~5 %. Each individual modification enhanced segmentation performance compared to the base model, with the full proposed architecture achieving the highest Dice scores. These results confirm that each component meaningfully contributes to the overall accuracy and robustness of the model. Overall, the model achieved Dice scores of 0.83 in FRDA patients and 0.85 in controls, which are comparable to or exceed those reported in prior work. For example, Chai et al. (Chai et al., 2022) reported a Dice score of 0.795 for whole DN segmentation using 88 subjects, while Shiraishi et al. (Shiraishi et al., 2025) achieved 0.897 (FRDA) and 0.891 (controls) with 355 subjects and QSM data. Beliveau et al. (Beliveau et al., 2021), using SWI and a similarly small dataset (50 controls), reported lower Dice scores of 0.77 and 0.78 for the left and right DN, respectively.

Our results align closely with those of Shiraishi et al. (Shiraishi et al., 2025), who used QSM but with nearly ten times more datasets. While QSM imaging provides higher visibility of DN regions, it requires additional post-processing steps. In contrast, T2*-weighted imaging offers broader clinical feasibility and more straightforward implementation, making it a practical choice for routine clinical use. In our segmentation

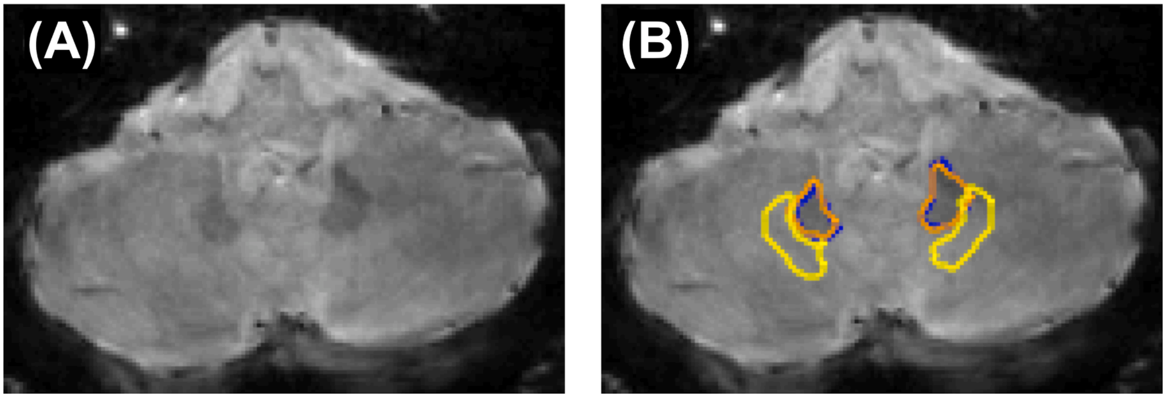

Fig. 2. **(A)** Original T2*-weighted MRI slice. **(B)** Segmentation of the dentate nuclei using manual annotation (blue), an atlas-based method (yellow) and our machine learning-based method (orange), illustrating the improved accuracy and delineation achieved by the proposed model.

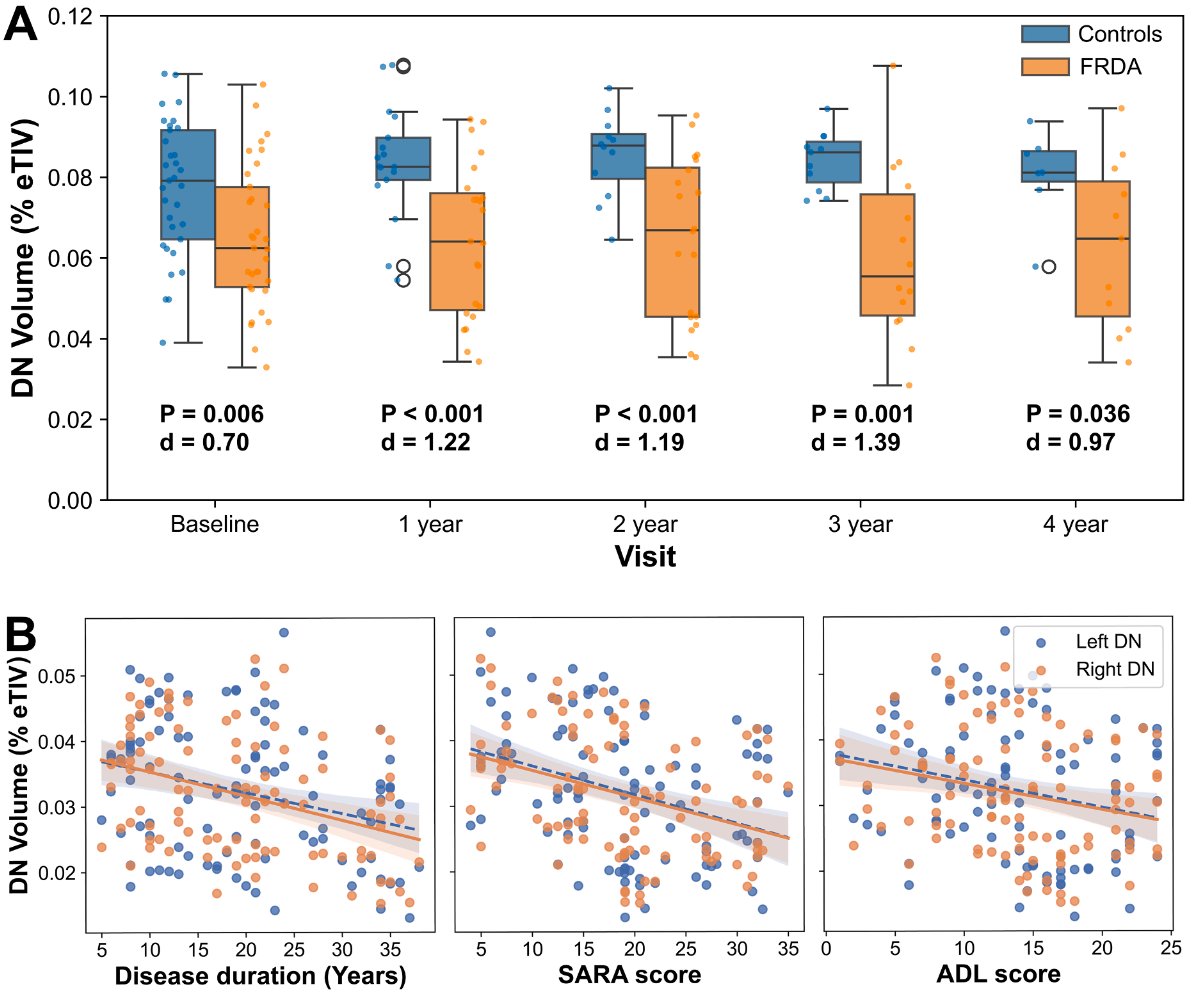

Fig. 3. **(A)** Box plots showing the distribution of dentate nuclei (DN) volumes expressed as percentage of the estimated total intracranial volume (%eTIV) for healthy controls and FRDA patients at baseline and follow-up visits over four years. P-values indicate statistical significance of between-group differences with Cohen's d as an effect size estimate. **(B)** Scatter plots of left and right DN volumes expressed as %eTIV, plotted against disease duration in years, SARA, and ADL scores. Solid regression lines indicate significant associations over time for the right DN at $p < 0.05$ based on linear mixed-effect modelling.

results, controls exhibited slightly higher Dice scores compared to FRDA patients for both left and right DN regions, reflecting the expected challenge of segmenting atrophic and structurally altered DN in FRDA. The ability to achieve state-of-the-art results with a small dataset, while maintaining high accuracy despite the structural abnormalities commonly associated with disease pathology, underscores the effectiveness of transfer learning in addressing data scarcity, a critical challenge in rare diseases like FRDA. Unlike prior studies that relied on high-work-intensive manual segmentation or atlas-based methods (Harding et al., 2021), our approach offers a more accessible, faster and automated method using clinical-standard available T2*-weighted imaging. Significantly improved results for our ML-based method compared to the classical atlas-based method underscore the advantage of data-driven segmentation in capturing subject-specific variability often missed by traditional atlas-based approaches.

Cross-sectional group comparisons revealed significant reductions in DN volumes in FRDA patients compared to controls, with medium to large effect sizes across baseline and follow-up visits. These findings are consistent with prior studies using QSM (Georgiou-Karistianis et al., 2025; Harding et al., 2016, 2024; Ward et al., 2019). Importantly, to our knowledge, this is the first study to report longitudinal data of DN volume changes over a follow-up time of four years. Although we did not find significant progression rates of DN atrophy with time, DN volumes declined with longer disease duration, ataxia severity measured by SARA, and functional impairment in ADL, which aligns with the slowly progressive nature of FRDA. A QSM study by Ward et al. (Ward et al., 2019) investigated iron concentration and volume changes in the DN after two years and reported no significant longitudinal atrophy rates compared to controls. Longitudinal change, however, was related to disease duration and ataxia severity (Ward et al., 2019). These findings in adult individuals with FRDA, comparable to our cohort, are in good agreement with our ML-based results and suggest that DN atrophy, which is already evident at early disease stages, may eventually reach a plateau over time. This notion is supported by recently published cross-sectional data from the first global, prospective imaging study in FRDA (TRACK-FA), which reported relatively constant DN atrophy over disease stages and absence of clinical correlations (Georgiou-Karistianis et al., 2025). Ongoing longitudinal analysis of the younger, particularly pediatric, TRACK-FA cohort may help to identify critical stages of DN vulnerability during disease progression. In addition, it is important to note that recent shape analysis of the DN in FRDA indicated that structural changes are not spatially uniform, hence, regional measures of DN substructures may be more sensitive measures of disease progression (Harding et al., 2016). Interestingly, in our study, correlations with clinical measures over time were observed for right DN volumes, with the majority of patients (88 %) being right-handed. While this may indicate an increased vulnerability of the preferentially used (ipsilateral) motor hemisphere of the cerebellum, and possibly of right-lateralized language functions, Ward et al. (Ward et al., 2019) reported an opposite pattern with more left-lateralized correlations. Further research is required to elucidate potential lateralization effects of DN structures in tracking disease progression in FRDA.

Despite its promising results, our study has several limitations. The sample size used for optimizing and validating the ML models was relatively small, and further validation with a larger dataset is needed to confirm generalizability. While T2*-weighted imaging is widely accessible, direct comparisons with SWI and QSM-based methods could enhance our approach and its applicability. Specifically, QSM provides a superior contrast of cerebellar nuclei, including the prominent dentate gyrification, and additionally enables quantification of iron levels. Although the latter is not possible with T2*-weighted imaging alone, histopathological studies reported normal total iron levels in the DN (Koeppen et al., 2007), questioning the relevance of DN iron concentration as a biomarker in FRDA. In vivo MRI studies using magnetic resonance relaxometry as a measure of iron content in DN have yielded conflicting results (Boddaert et al., 2007; Bonilha da Silva et al., 2014; Solbach et al., 2014; Waldvogel et al., 1999). In contrast, quantitative assessments using QSM indicate elevated DN iron level (Georgiou-Karistianis et al., 2025; Harding et al., 2016; Ward et al., 2019), but recent evidence from voxel-level QSM studies suggests spatially variable or circumscribed susceptibility increases that are more in accordance with post-mortem findings (Deistung et al., 2022; Harding et al., 2024). Another disadvantage of T2*-weighted images and SWI is that iron-rich regions appear enlarged, known as the blooming effect (Haacke et al., 1995). While this effect can modestly inflate absolute DN volumes, its impact on relative group differences and correlations with clinical measures is mitigated through moderately long echo times, carefully annotated training data, and standardized intensity normalization in our ML-based segmentation. We did not implement a specific correction for blooming, and its magnitude remains sequence- and context-dependent. Therefore, absolute volumes may be slightly overestimated, but the relative trends and associations reported here are unlikely to be substantially affected. Although the volumes measured in this study may be slightly overestimated, although they are in a similar range to those reported in other QSM studies (Deistung et al., 2022; Harding et al., 2024). Nevertheless, they will inherently vary from histological measures and should be seen as an estimation relative to control data. Finally, our longitudinal dataset contained missing data, as not all participants attended all five visits. To address this, we utilized linear mixed-effects modeling with restricted maximum likelihood estimation, allowing for inference even with incomplete datasets. Moreover, the interpretability of the deep learning models is limited, as we did not perform saliency or feature attribution analyses, which remain an important direction for future work.

Overall, this study demonstrates the feasibility and clinical relevance of an ML-based approach for segmenting the DN using commonly acquired T2*-weighted MRI, even with limited data availability. To address the challenge of data scarcity inherent to rare diseases such as Friedrich's ataxia, we adopted a transfer learning strategy involving pretraining, which avoids the need for large, manually annotated datasets. The pretrained model successfully generalized to the task of DN segmentation and enabled volumetric analysis across a larger cohort. Our findings revealed significant DN volume reductions in FRDA patients compared to healthy controls, with medium to large effect sizes. These volumetric changes were associated with longer disease duration and worsening clinical scores (SARA and ADL), highlighting the potential utility of DN volume as a neuroimaging marker for disease progression. By leveraging clinically accessible imaging modalities and robust ML techniques, this approach offers a scalable and reproducible framework for segmentation of the DN. Further evaluation on larger and independent datasets is required to assess its potential for providing insights into neurodegenerative changes associated with FRDA.

Author contributions

TS, ID, SR, and KR contributed to the conception and design of the study. TS, RD, LH, SR, KR, and ID contributed to the acquisition and analysis of the data. RA refined machine learning methodology. All authors contributed to the writing and editing of the manuscript and approved the final manuscript.

Financial disclosure

The authors have no financial disclosures relevant to this work. In the preceding 12 months: KR has received grants from the German Research Foundation, Friedreich's Ataxia Research Alliance (FARA), Interdisciplinary Center for Clinical Research within the faculty of Medicine at the RWTH Aachen University, Germany (OC2), German Federal Ministry of Education and Research (BMBF KP22- 106E) and honoraria for presentations or advisory boards from Biogen, Eisai, Lilly and Roche. ID received grants from the Friedreich's Ataxia Research Alliance (FARA) and Interdisciplinary Center for Clinical Research within the faculty of Medicine at the RWTH Aachen University, Germany (OC2). RD received grants from the Friedreich's Ataxia Research Alliance (FARA). JBS received grants from the German Research Foundation, Interdisciplinary Center for Clinical Research within the faculty of Medicine at the RWTH Aachen University, Germany, Biogen and Eisai and for presentations or advisory boards from Biogen, Reata, Eisai, Lilly, Roche, NovoNordisk. KMS is supported by the Deutsche Forschungsgemeinschaft (German Research Foundation; TRR 219; Project-ID 322,900,939 [C07]).

Funding

This project was supported by a grant from the Interdisciplinary Centre for Clinical Research (IZKF) within the faculty of Medicine at the RWTH Aachen University (OC2–2).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was funded by a grant from the Interdisciplinary Centre for Clinical Research (IZKF) within the faculty of Medicine at the RWTH Aachen University (OC2–2) and supported by the Brain Imaging Facility (BIF) of IZKF. Computations were performed with computing resources granted by RWTH Aachen University under project rwth1678. We thank Shahram Mirzazade for his valuable contribution by performing MR imaging. We thank the EFACTS consortium, all study participants and their families.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mlwa.2025.100790.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

Alshardan, A., Alruwais, N., Alqahtani, H., Alshuhail, A., Almukadi, W. S., & Sayed, A. (2024). Leveraging transfer learning-driven convolutional neural network-based semantic segmentation model for medical image analysis using MRI images. *Scientific Reports, 14*(1), Article 30549. https://doi.org/10.1038/s41598-024-81966-y

Azad, B., Azad, R., Eskandari, S., Bozorgpour, A., Kazerouni, A., Rekik, I., & Merhof, D. (2023). *Foundational models in medical imaging: A comprehensive survey and future vision* (No. arXiv:2310.18689). arXiv. https://doi.org/10.48550/arXiv.2310.18689.

Azad, R., Asadi-Aghbolaghi, M., Fathy, M., & Escalera, S. (2019). Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (pp. 406–415). https://doi.org/10.1109/ICCVW.2019.00052

Beliveau, V., Nørgaard, M., Birkl, C., Seppi, K., & Scherfler, C. (2021). Automated segmentation of deep brain nuclei using convolutional neural networks and susceptibility weighted imaging. *Human Brain Mapping, 42*(15), 4809–4822. https://doi.org/10.1002/hbm.25604

Bermudez Noguera, C., Bao, S., Petersen, K. J., Lopez, A. M., Reid, J., Plassard, A. J., Zald, D. H., Claassen, D. O., Dawant, B. M., & Landman, B. A. (2019). Using deep learning for a diffusion-based segmentation of the dentate nucleus and its benefits over atlas-based methods. *Journal of Medical Imaging, 6*(4), Article 044007. https://doi.org/10.1117/1.JMI.6.4.044007

Boddaert, N., Le Quan Sang, K. H., Rötig, A., Leroy-Willig, A., Gallet, S., Brunelle, F., Sidi, D., Thalabard, J.-C., Munnich, A., & Cabantchik, Z. I. (2007). Selective iron chelation in Friedreich ataxia: Biologic and clinical implications. *Blood, 110*(1), 401–408. https://doi.org/10.1182/blood-2006-12-065433

Bonilha da Silva, C., Bergo, F. P. G., D'Abreu, A., Cendes, F., Lopes-Cendes, I., & França, M. C., Jr (2014). Dentate nuclei T2 relaxometry is a reliable neuroimaging marker in Friedreich's ataxia. *European Journal of Neurology, 21*(8), 1131–1136. https://doi.org/10.1111/ene.12448

Bottou, L. (2014). From machine learning to machine reasoning. *Machine Learning, 94*(2), 133–149. https://doi.org/10.1007/s10994-013-5335-x

Brodley, C. E., Rebbapragada, U., Small, K., & Wallace, B. (2012). Challenges and opportunities in applied machine learning. *AI Magazine, 33*(1), Article 1. https://doi.org/10.1609/aimag.v33i1.2367

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research, 70*, 245–317. https://doi.org/10.1613/jair.1.12228

Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Cañizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., … Pandolfo, M. (1996). Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science, 271*(5254), 1423–1427. https://doi.org/10.1126/science.271.5254.1423

Chai, C., Wu, M., Wang, H., Cheng, Y., Zhang, S., Zhang, K., Shen, W., Liu, Z., & Xia, S. (2022). CAU-Net: A deep learning method for deep gray matter nuclei segmentation. *Frontiers in Neuroscience, 16*, Article 918623. https://doi.org/10.3389/fnins.2022.918623

Chavhan, G. B., Babyn, P. S., Thomas, B., Shroff, M. M., & Haacke, E. M. (2009). Principles, techniques, and applications of T2*-based MR imaging and its special applications. *Radiographics: A Review Publication of the Radiological Society of North America, Inc, 29*(5), 1433–1449. https://doi.org/10.1148/rg.295095034

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), *Medical image computing and computer-assisted intervention – miccai 2016* (pp. 424–432). Springer International Publishing. https://doi.org/10.1007/978-3-319-46723-8_49.

Deistung, A., Jäschke, D., Draganova, R., Pfaffenrot, V., Hulst, T., Steiner, K. M., Thieme, A., Giordano, I. A., Klockgether, T., Tunc, S., Münchau, A., Minnerop, M., Göricke, S. L., Reichenbach, J. R., & Timmann, D. (2022). Quantitative susceptibility mapping reveals alterations of dentate nuclei in common types of degenerative cerebellar ataxias. *Brain Communications, 4*(1), fcab306. https://doi.org/10.1093/braincomms/fcab306

Deistung, A., Stefanescu, M. R., Ernst, T. M., Schlamann, M., Ladd, M. E., Reichenbach, J. R., & Timmann, D. (2016). Structural and functional magnetic resonance imaging of the cerebellum: Considerations for assessing cerebellar ataxias. *The Cerebellum, 15*(1), 21–25. https://doi.org/10.1007/s12311-015-0738-9

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage, 46*(1), 39–46. https://doi.org/10.1016/j.neuroimage.2009.01.045

Diedrichsen, J., Maderwald, S., Küper, M., Thürling, M., Rabe, K., Gizewski, E. R., Ladd, M. E., & Timmann, D. (2011). Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *NeuroImage, 54*(3), 1786–1794. https://doi.org/10.1016/j.neuroimage.2010.10.035

Doan, N. T., Xivry, J. O.de, & Macq, B. (2010). Effect of inter-subject variation on the accuracy of atlas-based segmentation applied to human brain structures. *Medical Imaging 2010: Image Processing, 7623*, 570–579. https://doi.org/10.1117/12.845586

Dogan, I., Romanzetti, S., Didszun, C., Mirzazade, S., Timmann, D., Saft, C., Schöls, L., Synofzik, M., Giordano, I. A., Klockgether, T., Schulz, J. B., & Reetz, K. (2019). Structural characteristics of the central nervous system in Friedreich ataxia: An in vivo spinal cord and brain MRI study. *Journal of Neurology, Neurosurgery & Psychiatry, 90*(5), 615–617. https://doi.org/10.1136/jnnp-2018-318422

Georgiou-Karistianis, N., Corben, L. A., Lock, E. F., Bujalka, H., Adanyeguh, I., Corti, M., Deelchand, D. K., Delatycki, M. B., Dogan, I., Farmer, J., França, M. C., Jr., Gabay, A. S., Gaetz, W., Harding, I. H., Joers, J., Lax, M. A., Li, J., Lynch, D. R., Mareci, T. H., … Henry, P.-G. (2025). Neuroimaging biomarkers for Friedreich Ataxia: A cross-sectional analysis of the TRACK-FA study. *Annals of Neurology, 98*(2), 386–397. https://doi.org/10.1002/ana.27237

Gibbons, E., Hoffmann, M., Westhuyzen, J., Hodgson, A., Chick, B., & Last, A. (2023). Clinical evaluation of deep learning and atlas-based auto-segmentation for critical organs at risk in radiation therapy. *Journal of Medical Radiation Sciences, 70*(S2), 15–25. https://doi.org/10.1002/jmrs.618

Haacke, E. M., Lai, S., Yablonskiy, D. A., & Lin, W. (1995). In vivo validation of the bold mechanism: A review of signal changes in gradient echo functional MRI in the presence of flow. *International Journal of Imaging Systems and Technology, 6*(2–3), 153–163. https://doi.org/10.1002/ima.1850060204

Haacke, E. M., Xu, Y., Cheng, Y.-C. N., & Reichenbach, J. R. (2004). Susceptibility weighted imaging (SWI). *Magnetic Resonance in Medicine, 52*(3), 612–618. https://doi.org/10.1002/mrm.20198

Harding, I. H., Chopra, S., Arrigoni, F., Boesch, S., Brunetti, A., Cocozza, S., Corben, L. A., Deistung, A., Delatycki, M., Diciotti, S., Dogan, I., Evangelisti, S., França, M. C., Göricke, S. L., Georgiou-Karistianis, N., Gramegna, L. L., Henry, P., Hernandez-Castillo, C. R., Hutter, D., … Thompson, P. M. (2021). Brain structure and degeneration staging in Friedreich Ataxia: Magnetic Resonance Imaging volumetrics from the ENIGMA-Ataxia Working Group. *Annals of Neurology, 90*(4), 570–583. https://doi.org/10.1002/ana.26200

Harding, I. H., Nur Karim, M. I., Selvadurai, L. P., Corben, L. A., Delatycki, M. B., Monti, S., Saccà, F., Georgiou-Karistianis, N., Cocozza, S., & Egan, G. F (2024). Localized changes in dentate nucleus shape and magnetic susceptibility in Friedreich Ataxia. *Movement Disorders, 39*(7), 1109–1118. https://doi.org/10.1002/mds.29816

Harding, I. H., Raniga, P., Delatycki, M. B., Stagnitti, M. R., Corben, L. A., Storey, E., Georgiou-Karistianis, N., & Egan, G. F. (2016). Tissue atrophy and elevated iron concentration in the extrapyramidal motor system in Friedreich ataxia: The IMAGE-FRDA study. *Journal of Neurology, Neurosurgery & Psychiatry, 87*(11), 1261–1263. https://doi.org/10.1136/jnnp-2015-312665

He, N., Langley, J., Huddleston, D. E., Ling, H., Xu, H., Liu, C., Yan, F., & Hu, X. P. (2017). Improved Neuroimaging Atlas of the Dentate Nucleus. *Cerebellum, 16*(5–6), 951–956. https://doi.org/10.1007/s12311-017-0872-7

Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., & Reuter, M. (2020). FastSurfer—A fast and accurate deep learning based neuroimaging pipeline. *NeuroImage, 219*, Article 117012. https://doi.org/10.1016/j.neuroimage.2020.117012

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging, 22*(1), 69. https://doi.org/10.1186/s12880-022-00793-7

Koeppen, A. H., & Mazurkiewicz, J. E. (2013). Friedreich ataxia: Neuropathology revised. *Journal of Neuropathology and Experimental Neurology, 72*(2), 78–90. https://doi.org/10.1097/NEN.0b013e31827e5762

Koeppen, A. H., Michael, S. C., Knutson, M. D., Haile, D. J., Qian, J., Levi, S., Santambrogio, P., Garrick, M. D., & Lamarche, J. B. (2007). The dentate nucleus in Friedreich's ataxia: The role of iron-responsive proteins. *Acta Neuropathologica, 114*(2), 163–173. https://doi.org/10.1007/s00401-007-0220-y

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2014). *Deeply-Supervised Nets* (No. arXiv:1409.5185). arXiv. https://doi.org/10.48550/arXiv.1409.5185.

Liu, C., Li, W., Tong, K. A., Yeom, K. W., & Kuzminski, S. (2015). Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain. *Journal of Magnetic Resonance Imaging, 42*(1), 23–41. https://doi.org/10.1002/jmri.24768

Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. *Nature Communications, 15*(1), 654. https://doi.org/10.1038/s41467-024-44824-z

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565–571). https://doi.org/10.1109/3DV.2016.79

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). *Attention U-Net: Learning Where to Look for the Pancreas* (No. arXiv:1804.03999). arXiv. https://doi.org/10.48550/arXiv.1804.03999.

Ravanfar, P., Loi, S. M., Syeda, W. T., Van Rheenen, T. E., Bush, A. I., Desmond, P., Cropley, V. L., Lane, D. J. R., Opazo, C. M., Moffat, B. A., Velakoulis, D., & Pantelis, C. (2021). Systematic Review: Quantitative Susceptibility Mapping (QSM) of Brain Iron Profile in Neurodegenerative Diseases. *Frontiers in Neuroscience, 15*, Article 618435. https://doi.org/10.3389/fnins.2021.618435

Reetz, K., Dogan, I., Costa, A. S., Dafotakis, M., Fedosov, K., Giunti, P., Parkinson, M. H., Sweeney, M. G., Mariotti, C., Panzeri, M., Nanetti, L., Arpa, J., Sanz-Gallego, I., Durr, A., Charles, P., Boesch, S., Nachbauer, W., Klopstock, T., Karin, I., … Schulz, J. B. (2015). Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: A cross-sectional analysis of baseline data. *The Lancet Neurology, 14*(2), 174–182. https://doi.org/10.1016/S1474-4422(14)70321-7

Reetz, K., Dogan, I., Hilgers, R.-D., Giunti, P., Mariotti, C., Durr, A., Boesch, S., Klopstock, T., De Rivera, F. J. R., Schöls, L., Klockgether, T., Bürk, K., Rai, M., Pandolfo, M., Schulz, J. B., Nachbauer, W., Eigentler, A., Depondt, C., Benaich, S., … Sweeney, M. G. (2016). Progression characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS): A 2 year cohort study. *The*

Lancet Neurology, 15(13), 1346–1354. https://doi.org/10.1016/S1474-4422(16)30287-3

Reetz, K., Dogan, I., Hilgers, R.-D., Giunti, P., Parkinson, M. H., Mariotti, C., Nanetti, L., Durr, A., Ewenczyk, C., Boesch, S., Nachbauer, W., Klopstock, T., Stendel, C., Rodríguez De Rivera Garrido, F. J., Rummey, C., Schöls, L., Hayer, S. N., Klockgether, T., Giordano, I., … Vasco, G. (2021). Progression characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS): A 4-year cohort study. *The Lancet Neurology, 20*(5), 362–372. https://doi.org/10.1016/S1474-4422(21)00027-2

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – miccai 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

Schmitz-Hübsch, T., du Montcel, S. T., Baliko, L., Berciano, J., Boesch, S., Depondt, C., Giunti, P., Globas, C., Infante, J., Kang, J.-S., Kremer, B., Mariotti, C., Melegh, B., Pandolfo, M., Rakowicz, M., Ribai, P., Rola, R., Schöls, L., Szymanski, S., … Klockgether, T. (2006). Scale for the assessment and rating of ataxia. *Neurology, 66* (11), 1717–1720. https://doi.org/10.1212/01.wnl.0000219042.60538.92

Shafighfard, T., Asgarkhani, N., Kazemi, F., & Yoo, D.-Y. (2025). Transfer learning on stacked machine-learning model for predicting pull-out behavior of steel fibers from concrete. *Engineering Applications of Artificial Intelligence, 158*, Article 111533. https://doi.org/10.1016/j.engappai.2025.111533

Shiraishi, D. H., Saha, S., Adanyeguh, I. M., Cocozza, S., Corben, L. A., Deistung, A., Delatycki, M. B., Dogan, I., Gaetz, W., Georgiou-Karistianis, N., Graf, S., Grisoli, M., Henry, P.-G., Jarola, G. M., Joers, J. M., Langkammer, C., Lenglet, C., Li, J., Lobo, C. C., … Rezende, T. J. R. (2025). Automated Deep Learning-based Segmentation of the Dentate Nucleus Using Quantitative Susceptibility Mapping MRI. *Radiology: Artificial Intelligence.* , Article e240478. https://doi.org/10.1148/ryai.240478

Solbach, K., Kraff, O., Minnerop, M., Beck, A., Schöls, L., Gizewski, E. R., Ladd, M. E., & Timmann, D. (2014). Cerebellar pathology in Friedreich's ataxia: Atrophied dentate nuclei with normal iron content. *NeuroImage : Clinical, 6*, 93–99. https://doi.org/10.1016/j.nicl.2014.08.018

Subramony, S. H., May, W., Lynch, D., Gomez, C., Fischbeck, K., Hallett, M., Taylor, P., Wilson, R., & Ashizawa, T. (2005). Measuring Friedreich ataxia: Interrater reliability of a neurologic rating scale. *Neurology, 64*(7), 1261–1262. https://doi.org/10.1212/01.WNL.0000156802.15466.79

Waldvogel, D., Van Gelderen, P., & Hallett, M. (1999). Increased iron in the dentate nucleus of patients with Friedreich's ataxia. *Annals of Neurology, 46*(1), 123–125. https://doi.org/10.1002/1531-8249(199907)46:1%253C123::AID-ANA19%253E3.0.CO;2-H

Ward, P. G. D., Harding, I. H., Close, T. G., Corben, L. A., Delatycki, M. B., Storey, E., Georgiou-Karistianis, N., & Egan, G. F. (2019). Longitudinal evaluation of iron concentration and atrophy in the dentate nuclei in friedreich ataxia. *Movement Disorders, 34*(3), 335–343. https://doi.org/10.1002/mds.27606