Al-Nahrain Journal for Engineering Sciences

Al-Nahrain J. Eng. Sci. 29(2) 392-401, 2026
Original Article

# Standardizing Cervical Spine MRI for AI: A Mathematically Formalized Preprocessing and Harmonization Pipeline

**Noel John Veigas[1], Dasharathraj K Shetty[2], Shyamasunder Bhat N[3], Prithvishree Ravindra[4], and Praveen Shastry[5]●.**

*1,2,5Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India*
*3Department of Orthopedics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal - 576104, India.*
*4Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India*

●*These authors contributed equally to this work*

**DOI:** https://www.doi.org/10.29194/NJES.29020392

**Received:** March 4, 2026 **Revised:** April 26, 2026 **Accepted:** May 2, 2026 **Published:** June 20, 2026

## Abstract

The standardization and preprocessing decisions determine the performance, reproducibility and generalizability of Artificial Intelligence (AI) models in cervical spine Magnetic Resonance Imaging (MRI). In this research, the author suggests a six-step standardized preprocessing pipeline that can help with the reliable analysis of heterogeneous MRI data using AI. It is obtained by the pipeline based on recurrent methodological patterns found in the literature, focusing on the operations reported to stabilize image intensity distribution and reduce unwanted sources of variability. A systematic literature review was screened on 2,882 records, and 43 articles were included per the inclusion criteria. Due to the large variability of imaging procedures, preprocessing plans, and reported outcomes metrics, no quantitative meta-analysis was performed and a qualitative synthesis was done. Basic methods such as segmentation, denoising, intensity normalization, augmentation, bias-field correction were consistently reported to improve AI performance. Studies evaluating tasks such as normalization and automated segmentation reported accuracies of 94% to 99.95%. Moreover, less-specific harmonization methods like Combatting Batch Effects (ComBat), z-score normalization and histogram matching, and Generative Adversarial Network (GAN)-based ones were also often attributed with lower scanner- and site-based variations. Regardless of these improvements, the AI workflow remains unstable because of the irregularity of predefined parameters, unequal standards of acquisition, and lack of methodological reporting. In efforts to deal with such challenges, the current work proposes a methodical six-stage preprocessing model that consists of quality control, noise removal, intensity normalization, anatomical segmentation, harmonization and validation. The suggested workflow model offers a practical and transparent basis of clinically translatable AI usage in cervical spine MRI based on conventional mathematical equations and clear documentation.

**Keywords:** *Degenerative Cervical Myelopathy, ComBat Harmonization, Radiomics Reproducibility, Multi-Scanner, Standardization, Good Health and Well-being (Sustainable Development Goal- SDG 3), Industry, Innovation, Infrastructure (SDG 9)*

**Corresponding author:** Provide the corresponding author information and publisher here. E-mail address: raja.shetty@manipal.edu.in

## 1. Introduction

It is a consistent finding in the literature that preprocessing, such as noise reduction, data normalization, and data augmentation, fundamentally alters the performance of Artificial Intelligence models of cervical spine MRI. In cervical spine imaging, it was proven that preprocessing methods, namely intensity windowing and normalization, are key requirements that have a strong impact on the classification and detection accuracy of deep learning networks, including ResNet, DenseNet, and You Only Look Once (YOLO) [1]. Along with these methods, augmentation methods, including cropping and region-of-interest enhancement or cropping, have also been shown to improve model performance.

### 1.1 Addressing Data Heterogeneity and Scanner Variability

The variations in scanning parameters, including the dissimilarity of soft tissue signal intensity and non-uniformity of acquisitions, pose a serious statistical challenge to the analysis of spine images. Consequently, the methodologically sound methods of statistical harmonization of cervical spine MRI data are histogram matching, z-score normalization, and Combatting Batch Effects (ComBat) harmonization. These methods have been shown to minimize intra-scanner variations and improve downstream segmentation and automatic re-cropping policies have led to augmented datasets and high mean Intersection over Union (IoU) values [2].

### 1.2 Advanced Standardization and Clinical Detection

The advanced level of standardization techniques has also been developed to overcome these inconsistencies. Comparative analyses have suggested that Generative Adversarial Networks (GAN), specifically Conditional GANs (cGANs), are very stable in reducing image changes owing to change in scan settings, thus boosting performance consistency in tasks such as disease stage classification [3]. Regarding the detection of spinal fractures in the cervical area, it has been established that image normalization and augmentation at the first stage of preprocessing can provide a deep learning system with a high detection rate and minimal false-negative rates [4].

### 1.3 Generalizability and Inter-Observer Consistency

In addition to accuracy, a very challenging problem concerning the AI models being used is their generalizability, which is the case when using algorithms to operate on datasets other than their training data. To address the challenge of device-level heterogeneity without the need for extensive retraining, recent deep learning approaches have utilized generative adversarial networks to harmonize multi-vendor MRI data. For instance, Kim et al. [5] developed a framework integrating a Cycle-Consistent Generative Adversarial Network (CycleGAN)-based network with a segmentation network, employing a novel Radiomic Feature (RF) loss function to preserve texture information during harmonization. This approach successfully harmonized unpaired multi-vendor lumbar spine MRI datasets, achieving a Dice coefficient of 0.920 for intervertebral disc segmentation. Crucially, this method demonstrated that robust segmentation performance could be maintained on new datasets containing disease information without requiring additional manual annotation or model retraining, thereby significantly enhancing the generalizability of the AI workflow [5]. Finally, preprocessing must be a part of AI processes to make them less variable across human readers. A transformer-based deep learning model that has been improved using preprocessing displayed much higher interobserver consistency and better reporting efficiency in Degenerative Cervical Spondylosis (DCS) assessment than standard protocols [6].
Methodological evidence to inform the design of the proposed workflow was also obtained by considering related medical imaging fields in which preprocessing operations demonstrate modality-independent algorithmic behaviour (e.g., sensitivity to noise, normalization, and harmonization strategies). These fields of evidence are only applied to assist in making methodological arguments; all other clinical interpretations are based on cervical spine studies.
For this purpose, the research questions (RQs) of the study were as follows.

- RQ1: To what extent do preprocessing and standardization operations always warrant consideration in a standardized cervical spine MRI workflow to be used in AI applications?
- RQ 2: What do the differences in preprocessing operations and performance, as well as the reproducibility of AI workflows, depend on?
- RQ3: How can existing methodological evidence be operationalized to a standardized, mathematically formalized pre-processing workflow of cervical spine MRI?

Accordingly, the following research objectives (ROs) were derived:

- RO1: Preprocessing and standardization actions involved in the development of a cervical spine MRI preprocessing workflow to be used by AI.
- RO2: To determine the impact of variation in preprocessing operations on AI workflow performance.
- RO3: Obtain a single, consistent, and standardized preprocessing pipeline that enables consistent and multicenter-applicable cervical spine MRI analysis.

## 2. Methodology

### 2.1. Study Design

This study used a structured review of existing scientific literature to develop a standardized preprocessing and harmonization workflow for cervical spine MRI in AI applications. Prior studies were analyzed to identify preprocessing steps that were consistently useful, to identify common methodological patterns, and to combine them into a clear and reproducible framework for AI-based imaging analysis.

### 2.2. Data Sources and Search Strategy

A structured literature search was conducted in Scopus-indexed databases and Web of Science to identify methodological evidence relevant to the design of preprocessing and harmonization workflows for cervical spine MRI in AI applications. The search strategy combined terms related to the cervical spine, MRI, preprocessing, harmonization, and artificial intelligence. A representative search string was:
("cervical spine" OR "cervical cord" OR "cervical vertebrae") AND ("MRI" OR "magnetic resonance imaging") AND ("preprocessing" OR "normalization" OR "harmonization" OR "segmentation" OR "denoising") AND ("artificial intelligence" OR "machine learning" OR "deep learning").
Reference lists of included articles were reviewed to identify preprocessing steps applied to MRI data, harmonization techniques used to reduce inter-scanner variability, and workflow design choices adopted in prior cervical spine MRI AI studies. The study selection process is illustrated in Fig 1.

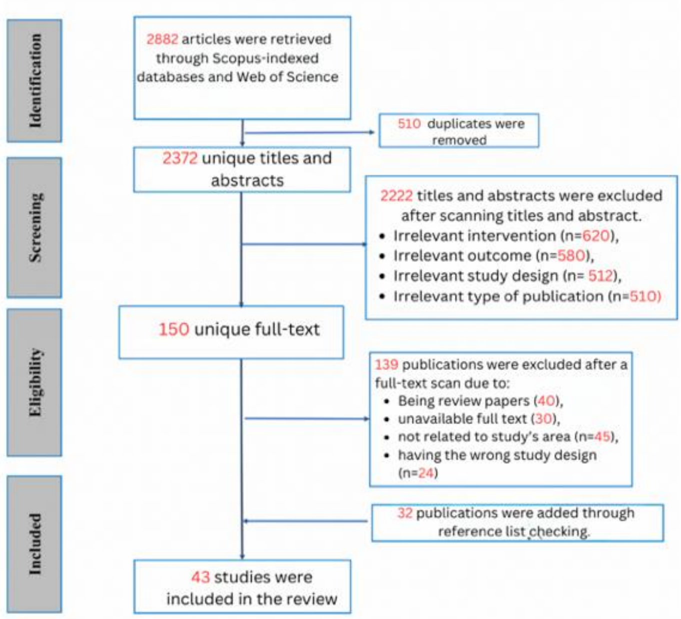


**Figure 1.** PRISMA flowchart summarizing the study selection process. (Source: Compiled by the authors)

### 2.3. Eligibility Criteria

Studies were screened using predefined inclusion and exclusion criteria to identify relevant methods, specifically preprocessing steps, harmonization techniques, and workflow design choices for developing a standardized cervical spine MRI preprocessing and harmonization workflow for AI applications.
Inclusion Criteria:

- Peer-reviewed journal articles or full conference papers.
- Studies involving cervical spine MRI implemented preprocessing or standardization operations for AI-based image analysis, including segmentation or feature extraction.
- Studies applying artificial intelligence, machine learning, deep learning, or radiomics methods.

Exclusion Criteria:
Studies were excluded if they were clearly unrelated to cervical spine MRI or to preprocessing and standardization operations relevant to AI workflows based on titles, abstracts, or accessible sections. Articles that did not report sufficient details on preprocessing steps, harmonization methods, or workflow structures were excluded, as were non–primary research papers. The review prioritized cross-sectional radiological imaging to ensure signal consistency, implicitly excluding methodologies derived from incompatible data structures unless they offered direct transferability to MRI physics.

## 2.4. Study Selection Process

All retrieved records were imported into Zotero for organization and deduplication. Titles, abstracts and other available sections were filtered to identify their relevance in cervical spine MRI preprocessing and standardization. Fig 2 gives a general description of this methodology.

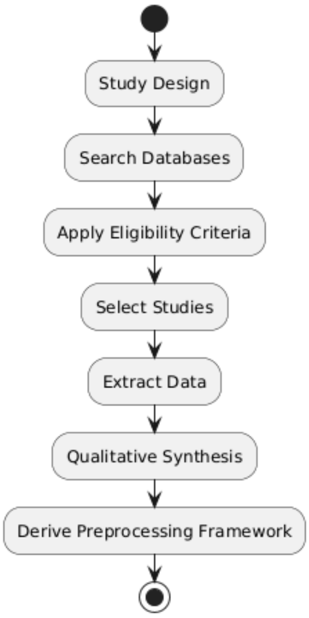


**Figure 2.** Overview of the methodology.
(Source: Compiled by the authors)

## 2.5. Data Extraction

The main study characteristics such as imaging parameters, preprocessing steps, harmonization methods, and AI methodological details were identified in every eligible article to aid in the synthesis.

## 2.6. Synthesis of Findings

The studies were vastly different in regards to imaging protocols, preprocessing procedures and outcome measures hence no quantitative meta-analysis could be performed. Instead, a qualitative synthesis was performed in a systematic manner in order to identify the trends that repeat and the influence they have on AI workflow resilience. It was synthesized using a three-level methodology.

- Frequency-Based Content Analysis: To quantify the frequency of utilization of some preprocessing operations.
- Thematic Synthesis: Gather these operations into functional groups (e.g., noise processing vs. harmonization).
- Stage-Gate Grouping: In order to have these functional categories put together in a standard or chronological process.

This methodology characterized the design of the architecture and the procedure of the proposed standardized preprocessing pipeline by mapping the most frequent and beneficial operations in a logical order. This approach defined the format and sequence of the intended standardized preprocessing pipeline by mapping most common and useful operations in a logical sequence.

## 2.7. Framework Derivation Approach

The standardized workflow suggested in preprocessing was created through generalization of methods that were observed or reoccurred in other studies. To construct a workflow, we

- We reviewed the sequencing of the preprocessing steps in other AI pipelines and found out sequencing that facilitated stable model performance.
- The preprocessing operations that were identified were of similar benefit to various datasets, scanners, and study designs.
- All the workflow stages were specified in standard mathematical formulations and generally accepted implementation practices, to make them understandable and consistent.

This led to a set of six steps of preprocessing workflow (Algorithm 1) that imbued scattered methodological practices into a single and reproducible pipeline that could be applied in AI applications to cervical spine MRI.

# 3.Preprocessing and Standardization Techniques

Preprocessing of cervical spine MRI signal processing takes into consideration four functional stages: noise suppression, intensity normalization, anatomical segmentation, and cross-site harmonization. Although certain applications may vary, the mathematical objective is the same: stabilize the statistical distribution of the image data prior to being input into an artificial intelligence algorithm.

## 3.1 Noise Suppression and Image Quality

Noise suppression focuses on image quality. MRI acquisition random noise usually has Rician or Gaussian distributions and may blur the delicate anatomical boundaries of the cervical cord. The overall aim of noise reduction is to minimize these high-frequency artifacts so that the Signal-to-Noise Ratio (SNR) is increased without losing acute edges to aid in defining vertebral boundaries.

## 3.2 Intensity Normalization and Signal Scaling

Signal scaling is performed by intensity normalization. Unlike in Computed Tomography (CT), the signal intensities on MRI are relative. Consequently, the two scans may also have vastly different pixel values for the same tissue type. Normalization is used to address this by ensuring that the raw pixel values are mapped onto

**Table 1.** An overview of the standardization and preprocessing methods. Source: Compiled by the authors

| Technique | Description | Category |
|---|---|---|
| Image Normalization [4, 7] | Adjusts pixel intensity values to a standard range | Preprocessing |
| Image Augmentation [4, [8], [9] | Increases dataset size through flipping, rotation, scaling | Preprocessing |
| Histogram Equalization. [9] | Enhances image contrast | Preprocessing |
| Skull Stripping and Slicing [10] | Removes non-brain tissues and segments images into slices | Preprocessing |
| GAN-based Standardization [3] | Uses cGAN and cycleGAN to standardise images against acquisition variations | Standardization |
| Z-Score Standardization[3] | Converts pixel values to a standard normal distribution | Standardization |
| ComBat Harmonization [3] | Adjusts for batch effects in multi-center studies | Standardization |
| Pathology-Guided Segmentation [2], [8], [11] | Segments key anatomical structures and pathological regions | Preprocessing |
| Manual Annotation [11, 12] | Expert radiologists provide ground truth labels | Preprocessing |
| ROI Detection [8], [11], [13] | Uses CNNs to detect and crop regions of interest | Preprocessing |
| Spinal Cord Toolbox (SCT) | Automated segmentation, vertebral labeling, and metric extraction | Preprocessing |

## 4.Influence of Preprocessing Variations on AI Model Performance

The effects here are associated with the same preprocessing steps in the workflow pipeline, and how the change in each step affects the model performance. There are four broad categories of operations that are commonly used in preprocessing in imaging to prepare data to be reliably modelled by AI. Fig 3 shows the four key types of preprocessing operations that are typically used in imaging processes

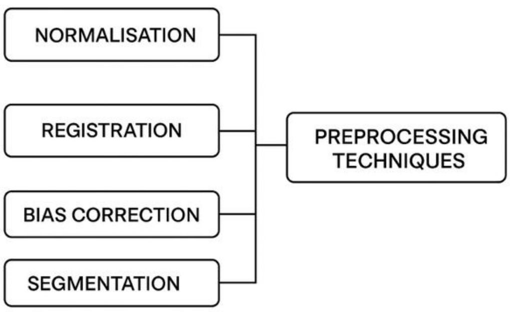


**Figure 3.** A high-level overview of the four primary preprocessing operations typically used in imaging workflows. (Source: Compiled by the authors)

### 4.1 Normalization Strategies and Signal Integrity

The accuracy, reproducibility and reliability of the AI models are highly influenced by the application of various preprocessing methods prior to the model training, including normalization, registration, bias correction and segmentation. The data must be normalized, especially to enable them to be consistent and comparable before they are subjected to computational analysis. Within the framework of cervical spine studies, the standardization of preprocessing processes is becoming known as a key to model reproducibility. The authors have shown that automated pipelines that combine processes like anatomically constrained smoothing and uniform mesh generation are necessary in ensuring the biofidelity of downstream computational models [14]. In contrast to generic normalization methods used in other applications, domain-specific MRI pipelines are aimed at stabilizing the geometric and signal properties of vertebrae and intervertebral discs, which allows the analysis of large cohorts of patients to be robust [15].

However, certain normalization techniques, including image intensity normalization, can decrease model accuracy and limit the capability of the model to identify subtle yet diagnostically important image features by decreasing the signal variance [16]. In the case of cervical spine MRI, standard processing structures have been necessary to ensure signal integrity in heterogeneous datasets. As an illustration, the Spinal Cord Toolbox (SCT) offers an integrative model, which uses MRI templates and atlases to spatially normalize spine data [17]. This method can be used to cross-validate and allows the accurate measurement of multi-parametric MRI measurements in multi-center studies by automating processes like motion correction and registration to common templates, overcoming the non-standardized nature of non-standardized workflows [17]. Various domains have been associated with better performance of Convolutional Neural Network (CNN) with normalization strategies (e.g., min-max scaling). In other reports, such as, isolated experiments have reported accuracy improvements of up to 94% and Mean Squared Error (MSE) of as low as 0.013 with internal evaluation, although these findings are highly dependent on the data set composition, preprocessing options and validation design [18].

### 4.2 Image Registration and Anatomical Alignment

It is not clear whether atlas registration is required when deep-learning models are used. There are studies that have indicated that it does not substantially enhance the performance of deep-learning models [16]. Registration with Different software tools and brain mapping methods (cortical parcellation methods) can have a strong impact on neurobiological outcomes and individual prediction tasks. This demonstrates that reproducibility is ensured by the necessity of the constant registration techniques [19]. Registration methods that may be applied to force anatomical structure into compliance can be useful in making datasets more homogenous, the impact of which on robustness may depend on datasets and applications, and whose performance may differ by up to 64 percentage points across preprocessing conditions [20].

### 4.3 Bias Correction and Generalizability Challenges

Bias field correction is necessary in cervical spine imaging to reduce the occurrence of non-uniformities in the intensity due to inhomogeneous magnetic field and differences in coil sensitivity.

Automated bias correction is included in standardized frameworks, including the Spinal Cord Toolbox (SCT), to provide consistent signal power across the vertebral column, which is important in the accurate measurement of multi-parametric MRI measures and the effective downstream segmentation [17]. Bias-field correction has little impact on model performance when medical imaging is used [16]. The role of bias correction in improving robustness is limited as its benefits are dependent on the dataset and may not work well in different situations, which lacks generalizability [16].

### 4.4 Segmentation and Regional Feature Extraction :

U-Net and Mask Region-based Convolutional Neural Network (Mask R-CNN) segmentation models demonstrated better results than state-of-the-art models [21]. Nevertheless, these tests are usually associated with corresponding datasets with the importance of cautious validation prior to making generalizability assumptions [21]. To perform operations such as feature extraction, model training, and diagnosis, it is important to consider regions of interest with proper segmentation. Because preprocessing variations can result in performance differences of up to 64 percentage points between configurations, segmentation techniques are essential for reproducibility [19], [20]. The development of deep learning architectures has greatly contributed to the accuracy of spinal imaging in anatomy segmentation. Most current research on lumbar spine MRI has shown that state-of-the-art models, such as MultiResUNet, may be used to attain remarkably high accuracy in segmenting vital organs such as the dural sac. MultiResUNet achieved an accuracy of 99.95% and a recall of 0.9989 in automated tasks of cross-sectional area Dural Sac Cross-sectional Area (DSCA) quantification on external validation tasks [22]. While MultiResUNet achieved an accuracy of 99.95% in lumbar spine applications [22], this serves as an optimistic proxy for cervical spine performance, which is anatomically smaller and more susceptible to motion artifacts. High-fidelity segmentation is not only necessary to decrease the radiological load burden, but it is also necessary to guarantee reproducibility that ensures uniform clinical diagnosis.

### 4.5 Strategic Pipeline Selection and Reproducibility:

The accuracy and reproducibility of AI models may be affected by the unpredictability introduced by choosing a preprocessing pipeline that combines multiple techniques. Testing and comparison of preprocessing pipelines must be performed carefully to reduce these effects [19] , [20]. Preprocessing steps should match the specific application and dataset to improve model performance, that is, different setups of normalization and resampling methods worked best for different clinical datasets [20]. Preprocessing choices have a strong impact on the performance of the AI model and careful selection with proper tuning is important for obtaining accurate, reproducible and reliable results. Since preprocessing tasks (e.g. normalization, noise reduction, harmonization, etc.) are performed based on image statistics and not anatomical position, information in other imaging fields provides transferable information regarding the stability and robustness of the algorithms.

## 5.Key Limitations, Challenges and Recommendations Identified in the Literature

### 5.1 Limitations in Data Acquisition and Reporting:

The MRI machines in various centers might be varied. The scanning parameters of the centers also decrease the generality of the AI-based diagnosis models [23], [24]. The scans may vary as a result of the above issues leading to poor image quality and diagnosis. The algorithms applied and the quality of the dataset are not clearly reported in many of the existing studies, which restricts their reproducibility and clinical utility [25]. The significant absence of standard procedures in cervical spine MRI results in variations in the imaging procedures and reporting [26], [27]. The open-access datasets are usually poorly documented and have uneven data formats that complicate the preprocessing and feature extraction process [28].

### 5.2 Challenges regarding Generalizability and Multi-Center Validation:

The models of AI tend to not perform well at working on external data, including data of other hospitals, scanners, or populations that are not similar to the training data, primarily due to device-level variations [5], [29]. If the data bias is low, the performance of the AI model is more reliable, but this remains a constant challenge [25]. Differences in radiomics features across multiple centers can change downstream model performance, which can make harmonization techniques necessary [23]. Advanced MRI techniques require large multicenter studies for validation to prove their effectiveness and support clinical use [30], thus improving generalizability.

### 5.3 Evidence-Based Recommendations for Standardization:

Consistency across multi-center studies is ensured by the development and application of standardized protocols and image processing techniques [24], [26], [27]. Self-supervised harmonization techniques are used to improve model generality, minimize inter-center disparities, and match image styles across various centers [23]. Reproducibility and clinical utility increase when algorithm design and dataset quality are made more transparent [25]. It is better to have a full-fledged AI-assisted clinical decision support system to provide effective and consistent assistance [25] throughout cervical spine surgery. In addition, ensuring that MRI tests meet the minimum requirements to be certified as quality could facilitate the standardization of imaging procedures [31].

## 6. Discussion

Findings of this research indicate that preprocessing and standardization have a strong influence on the efficacy, reproducibility, and generalizability of the AI models in cervical spine MRI. The consistency of the input data is what makes the results of the diagnosis reliable. Although the standardization methods, including histogram matching, z-score, ComBat, and GAN-based harmonization, reduced scanner-related variability and expanded the multi-center applicability, the preprocessing methods, including noise reduction, normalization, augmentation, segmentation, and annotation, enhanced the image quality and model accuracy.

Normalization improved signal quality, and the robustness was increased with the help of segmentation models like Mask R-CNN and U-Net. Registration and bias correction however had different impacts on accuracy. The choice of preprocessing pipeline and tuning should be according to specific applications, since the inconsistent setups yield inconsistent results. Issues that persist include inter-center variability, poor quality of data, and black box algorithms. Open data, standardized protocols and quality control can be important in enhancing reproducibility and trust. In general, to develop credible and clinically viable AI systems in cervical spine MRI, preprocessing and standardization are necessary. The performance measures mentioned in non-cervical areas were mentioned only to demonstrate the behavior of the methodology and cannot be regarded as clinically applicable to the cervical spine MRI.

The suggested six-step workflow, which is represented in a high-level procedural representation, provides a systematic basis of reproducible preprocessing in cervical spine MRI. Since preprocessing tasks (normalization, noise removals, and harmonization) are mathematically modality-agnostic, findings in broader imaging areas can be informative of algorithmic stability, even in a different clinical setting.

## 7. Recommended Framework for Standardized Cervical Spine MRI Preprocessing

A standardized preprocessing framework to support cervical spine MRI was suggested (Fig 4). The framework will facilitate uniform image acquisition, impose strict preprocessing and harmonization methods, and allow a clear documentation, thus creating a strong basis of multi-center and clinically validated AI applications.

To maintain the integrity of the data and maximize the model performance, the framework is structured as a sequential pipeline, where each step is based on the one preceding it. Variability is addressed at its source through quality control and acquisition standardization, followed by noise reduction and intensity normalization to stabilize image characteristics. Once intra-scan consistency is achieved, harmonization aligns data across scanners and centers, and anatomical segmentation enables the precise isolation of relevant structures. The workflow concludes with validation and documentation to ensure reproducibility and methodological transparency. Through this staged transformation, raw MRI data are progressively converted into harmonized inputs suitable for AI analysis, with each step designed to enhance, rather than distort, the integrity of the data. Table 2 provides mathematical expressions and examples corresponding to each step of the workflow pipeline, with formulas adapted from the literature, where indicated. The mathematical formulations in Table 2 were adapted based on the literature for the analysis of medical images. In cases where simple measures (e.g. SNR, MSE, or Dice coefficients) or algorithms (e.g., U-Net or ComBat) are used, the expressions are retained in the definitions provided in the original studies. The formulae have been modified in situations where decision logic or details of implementing the study, such as the quality control thresholding in Step 1 or the normalization limits in Step 3, so that the sense of the general logic of the proposed pipeline is still conveyed, but the theoretical foundation of the original works is retained.

| Step | Mathematical Expression | Comment | Notation Explanation | Sample Python Code |
|---|---|---|---|---|
| 1. Quality Control (QC) | $Qs = 1 if (SNRs \geq \tau) AND (Seqs = Ptgt)$<br>$Qs = 0\ otherwise$<br>[32] | Accept scan only if SNR is sufficient AND protocol matches target. | s: MRI scan index<br>$Q_s$: QC indicator (1 = pass, 0 = fail)<br>$\mathrm{SNR}_s$: Signal-to-noise ratio of scan $s$<br>$\tau$: Minimum acceptable SNR threshold<br>$\mathrm{Seq}_s$: Acquisition protocol of scan $s$<br>$P_{tgt}$: Target MRI protocol<br>$\wedge$: Logical AND (both conditions required) | import numpy as np<br>def qc_pass_simple(img, snr_thresh=10):<br>img = img[np.isfinite(img)]<br>bg = img[img < np.percentile(img, 20)]<br>fg = img[img > np.percentile(img, 80)]<br>return int(np.mean(fg) / np.std(bg) >= snr_thresh) |
| 2. Noise Reduction | $I_{denoised} = I \otimes K$<br>[33] | Apply filter to reduce noise. | $I$: input image; $K$: smoothing kernel; $\otimes$: convolution. | from scipy.ndimage import gaussian_filter<br>def denoise_gaussian(img, sigma=1.0):<br>return gaussian_filter(img, sigma=sigma)<br>I_denoised = denoise_gaussian(mri_img) |
| 3 A. Intensity Normalization (z-score) | $I_{norm}(x) = \frac{I(x) - \mu}{\sigma}$<br>[34] | Standardise intensity to zero mean, unit variance. | $x$: voxel; $\mu$: mean intensity; $\sigma$: std dev of intensities. | import numpy as np<br>def zscore_norm(img):<br>mu = img.mean()<br>sigma = img.std() or 1e-6<br>return (img - mu) / sigma<br>I_norm = zscore_norm(I_denoised) |
| 3 B. Intensity Normalization (min–max) | $I_{norm}(x) = \frac{I(x) - I_{min}}{I_{max} - I_{min}}$<br>[34] | Rescale intensity to [0, 1]. | $I_{min}, I_{max}$: min and max voxel intensity in image. | import numpy as np<br>def minmax_norm(img):<br>imin, imax = img.min(), img.max()<br>denom = (imax - imin) or 1e-6<br>return (img - imin) / denom<br>I_norm = minmax_norm(I_denoised) |
| 4 A. Segmentation (DL model) | $M(x) = f_\theta(I(x))$[35] | Model produces segmentation mask for each voxel. | $M(x)$: mask; $f_\theta$: network (e.g., U-Net) with parameters $\theta$. | import torch<br>with torch.no_grad():<br>logits = model(img_tensor)<br>probs = torch.sigmoid(logits)<br>M = (probs > 0.5).float()<br>M_np = M.squeeze().cpu().numpy() |

| | | | | |
|---|---|---|---|---|
| 4 B. ROI Cropping | $I_{\mathrm{ROI}} = I \cdot M$ [36] | Keep only segmented regions of interest. | $I_{\mathrm{ROI}}$: ROI image; $M$: binary mask; ·: element-wise product. | import numpy as np<br>I_roi = I_norm * M_np |
| 5 A. Harmonization (ComBat – abstract form) | $Y_{ij}^{\mathrm{ComBat}} = \dfrac{Y_{ij} - \hat{\alpha} - X_{ij}\hat{\beta} - \gamma_i^*}{\delta_i^*} + \hat{\alpha} + X_{ij}\hat{\beta}$ [37] | Adjust scanner/site effects while preserving biological signal. | where $Y_{ij}^{\mathrm{ComBat}}$ is the harmonized data for subject *j* at site *i*; $Y_{ij}$ is the original data; $\hat{\alpha}$ is the global intercept; $X_{ij}$ represents biological covariates; $\hat{\beta}$ are their regression coefficients; $\gamma_i^*$ and $\delta_i^*$ ar e the Empirical Bayes estimates of the additive (location) and multiplicative (scale) site effects, respectively. | from neurocombat_sklearn import CombatModel<br>combat = CombatModel()<br>Y_harmonized = combat.fit_transform(<br>features,<br>batch=batch,<br>covars=covariates<br>) |
| 5 B. Harmonization (Histogram Matching) | $I_{\mathrm{harm}}(x) = H_t^{-1}\left(H_s(I(x))\right)$ [38] | Match intensity histogram of source to target. | $H_s$: CDF of source; $H_t$: CDF of target; $H_t^{-1}$: inverse CDF. | from skimage.exposure import match_histograms<br>I_harm = match_histograms(I_norm, target_img) |
| 5 C. Harmonization (GAN-based) | $I_{\mathrm{harm}} = G(I)$ [39] | Generator maps source-domain image to target-domain style. | $G$: trained generator (cGAN/cycleGAN). | import torch<br>with torch.no_grad():<br>inp = torch.from_numpy(I_norm[None, None]).float().to(device)<br>I_harm_t = G(inp)<br>I_harm = I_harm_t.squeeze().cpu().numpy() |
| 6 A. Validation – Accuracy | $Acc = \dfrac{TP + TN}{TP + TN + FP + FN}$ [40] | Standard classification accuracy metric. | $TP, TN, FP, FN$: confusion matrix components. | import numpy as np<br>def accuracy(y_true, y_pred):<br>y_true, y_pred = np.asarray(y_true), np.asarray(y_pred)<br>return (y_true == y_pred).mean()<br>acc = accuracy(y_true, y_pred) |
| 6 B. Validation – MSE / Dice | $MSE\colon (1/\lvert\Omega\rvert)\Sigma_x\left(I_pred(x) - I_true(x)\right)^2$ *[41]*<br>Dice $= \dfrac{2\sum_x P(x)G(x)}{\sum_x P(x) + \sum_x G(x) + \varepsilon}$. [42] | Quantifies voxel-wise reconstruction or prediction error (MSE) and overlap between predicted and true segmentation masks (Dice). Quantifies segmentation overlap between predicted and ground-truth masks | Ω: voxel set; \|Ω\|: voxel count; $\Sigma_x$: sum over voxels; I_pred, I_true: predicted & true intensities; P, G: predicted & true masks; ε: stability constant. P(x): predicted mask; G(x): ground truth; ε: small stability constant | import numpy as np<br>def mse(I_pred, I_true):<br>return np.mean((I_pred.astype(np.float32) - I_true.astype(np.float32))**2)<br>def dice(P, G, eps=1e-6):<br>P = P.astype(np.float32)<br>G = G.astype(np.float32)<br>intersection = np.sum(P * G)<br>return (2 * intersection) / (np.sum(P) + np.sum(G) + eps) |
| 6 C. Documentation / Metadata | $\mathcal{P} = p, m, f, d$ [43] | Store pipeline configuration for reproducibility. | $\mathcal{P}$: set of all preprocessing and model parameters.<br>p: preprocessing parameters<br>m: model settings<br>f: filters applied<br>d: dataset identifiers | import json<br>config = {<br>"qc_snr_thresh": 10,<br>"denoise": {"method": "gaussian", "sigma": 1.0},<br>"norm": "zscore",<br>"model": "unet_v1",<br>"harmonization": "hist_match"}<br>with open("pipeline_config.json", "w") as f:<br>json.dump(config, f, indent=2) |

### 7.1 Overview of the Standardized Preprocessing Workflow:

Based on the synthesized literature, the proposed preprocessing pipeline can be expressed as a structured six-step workflow consisting of the following steps:

(1) Quality control and standardized acquisition
(2) Noise reduction
(3) Intensity normalization
(4) Anatomical segmentation
(5) Harmonization across scanners and centers
(6) Validation and documentation

One stage of the workflow is based on the earlier stage, which gradually transforms raw MRI data into harmonized artificial intelligence-ready inputs. To provide a clear categorization of the various preprocessing strategies reported in the literature, these operations were arranged in a condensed six-stage workflow. The structure demonstrates common methodological trends in the literature and portrays the methods typically used in preprocessing cervical spine MRI. In a high-level overview, this workflow is represented in Algorithm 1, in which the most important operations, including noise reduction, normalization, segmentation, and harmonization, are organized into a reproducible and transparent preprocessing pathway.

Algorithm 1. Standardized Preprocessing Pipeline for Cervical Spine MRI:

1: Input: Raw MRI scan I
2: Perform quality control: Verify SNR thresholds AND acquisition protocol (T1/T2/STIR)
3: If scan fails QC → discard or flag for manual review
4: Apply noise reduction filter to obtain I_denoised
5: Normalize intensity values: I_norm = Normalize (I_denoised)
6: Generate anatomical segmentation mask M = fθ (I_norm)
7: Extract region of interest: I_ROI = I_norm · M
8: Apply harmonization method (ComBat, histogram matching, or GAN)
9: Perform internal and external validation of processed outputs
10: Document all preprocessing parameters and transformations
11: Output: Harmonized, AI-ready MRI dataset

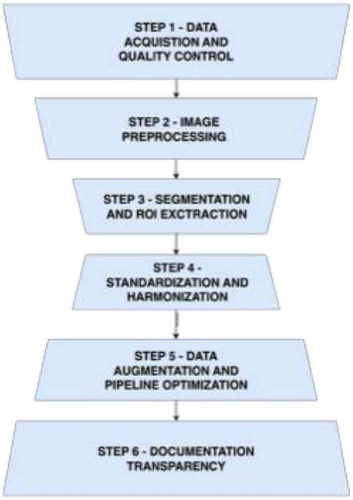


**Figure 4.** Proposed framework for preprocessing and standardization of cervical spine MRI data for AI applications. (Source: Compiled by the authors)

## 8. Limitations and future directions

Major variations in MRI protocols, sequences, and scanner equipment across research studies make direct comparisons impossible and reduce the ability to replicate. Complete reporting on preprocessing procedures, data nature, and measurement of evaluation further limits fair evaluation, as it did not allow systematic evaluation of the impact of the factors of acquisition on the effectiveness of preprocessing. The majority of the studies were based on small or single-center data with little outer validation of the generalizability of the reported AI performance.

Literature screening was conducted using titles, abstracts, and available sections, which may have introduced selection bias and resulted in the omission of relevant studies with insufficiently detailed methodological reporting. In addition, the available evidence may be influenced by publication bias, with a disproportionate representation of studies reporting positive outcomes. The limitations of standardization and reproducibility are still due to limited access to the tools, poor documentation, and inconsistency in implementing it across research groups.

Future research ought to focus on the use of open-source benchmarks, multi-center datasets, including complete metadata, comparisons between preprocessing strategies, and automated quality control and harmonization procedures. Strict methodology, interdisciplinary collaboration, and transparent reporting procedures are needed to ensure credible clinical translation.

## 9. Conclusion

This paper suggested a standard preprocessing and standardization workflow of cervical spine MRI in artificial intelligence (AI) applications, which is a major methodological obstacle to reproducibility and cross-center implementation. Through a systematic review of methodological evidence of previous studies, fundamental preprocessing procedures, such as noise reduction, intensity normalization, augmentation, anatomical segmentation and harmonization, were discovered and integrated into a coherent six-step workflow that aims to stabilize AI performance and facilitate generalizable model development.

The objectives of the research in terms of specificity were achieved as (i) the identification and establishment of preprocessing and standardization operations that will always be applicable to robust AI processes, (ii) the evaluation and subsequent establishment of the impact of altering such operations on model performance, reproducibility and generalizability, and (iii) operationalization's of such results into a single, mathematically formalized preprocessing pipeline applicable across centers.

The results highlight the fact that the quality of AI performance is highly contingent upon preprocessing decisions and that the performance metrics that are provided in single studies are frequently highly task- and dataset-specific, which restricts external validity. Although the suggested workflow offers an evidence-based basis of standardized preprocessing, it is a literature-based framework and has to be tested on large, heterogeneous, and multi-center datasets. To transform cervical spine MRI AI systems into clinically reliable and reproducible tools, the further evolution of open-source implementations, transparent reporting standards, and strict external evaluation will be required.

## 10. Acknowledgment

"To support academic integrity and clarity, this manuscript was prepared with the assistance of AI-based tools used for literature exploration, language refinement, and editorial support. Citation management and paraphrasing assistance were facilitated using Quillbot, whereas Grammarly and Paperpal were employed to identify grammatical, punctuation, and stylistic issues. Scopus AI was used as a literature discovery tool to support the efficient identification and synthesis of relevant studies, and Scholarcy assisted in summarizing and organizing prior research. Similarity screening was conducted using Turnitin. All scientific interpretations, analyses, and conclusions are the responsibility of the authors."

## 11. Declaration of Competing Interests

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

## 12. Funding Declaration

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

## 13. Ethical Approval

The study was conducted in accordance with the institutional guidelines for ethical research.

## References

[1] M. Yaseen, M. Ali, S. Ali, A. Hussain, M.-I. Joo, and H.-C. Kim, "Cervical spine fracture detection and classification using Two-Stage Deep Learning methodology," IEEE Access, vol. 12, pp. 72131–72142, (2024). https://doi.org/10.1109/access.2024.3398061

[2] Q. Zhang, X. Chen, Z. He, L. Wu, K. Wang, J. Sun, and H. Shen, "Pathology-guided AI system for accurate segmentation and diagnosis of cervical spondylosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 30, no. 2, pp. 1216–1229, Feb. (2026). https://doi.org/10.1109/JBHI.2025.3598469

[3] S. Niyoteka, R. Rouhi, P.-A. Laurent, S. Achkar, A. Carré, S. Diffetocq, C. Balleyguier, C. Chargari, E. Deutsch, and C. Robert, "GAN-based standardization of MR images: A promising approach for the development of multicenter radiomic models," *Physics in Medicine & Biology*, vol. 70, no. 18, (2025). https://doi.org/10.1088/1361-6560/adf2f4

[4] M. L. Kumar, P. Sampath, P. V. V. S. Srinivas, D. Dommeti, and S. R. Nallapati, "Leveraging Deep Learning for Accurate Detection and Precise Localization of Vertebral Fractures in Medical Imaging," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, (2023). https://doi.org/10.1109/ICOSEC58147.2023.10275972

[5] C. Kim, S.-M. Park, S. Lee, and D. Lee, "A deep learning harmonization of multi-vendor MRI for robust intervertebral disc segmentation," *IEEE Access*, vol. 12, pp. 19482 - 19499, (2024). https://doi.org/10.1109/ACCESS.2024.3360272

[6] A. Lee *et al.*, "Using deep learning to enhance reporting efficiency and accuracy in degenerative cervical spine MRI," *The Spine Journal*, vol. 25, no. 9, pp. 1942–1950, (2025). https://doi.org/10.1016/j.spinee.2025.03.009

[7] M. Singh, U. Tripathi, K. K. Patel, K. Mohit, and S. Pathak, "An efficient deep learning based approach for automated identification of cervical vertebrae fracture as a clinical support aid," *Scientific Reports*, vol. 15, Art. no. 25651, (2025). https://doi.org/10.1038/s41598-025-10448-6.

[8] E. A. Tarakçı *et al.*, "Novel preprocessing-based sequence for comparative MR cervical lymph node segmentation," *Journal of Clinical Medicine*, vol. 14, no. 6, p. 1802, (2025). http://doi.org/10.3390/jcm14061802

[9] J. Park, J. Yang, S. Park, and J. Kim, "Deep learning-based approaches for classifying foraminal stenosis using cervical spine radiographs," *Electronics*, vol. 12, no. 1, p. 195, (2023). http://doi.org/10.3390/electronics12010195

[10] V. Viswan, F. Hajamohideen, K. Subramanian, N. Shaffi, and M. Mahmud, "Enhancing insights: Unravelling the potential of preprocessing MRI for artificial intelligence based Alzheimer's disease classification," in *Machine Learning Models and Architectures for Biomedical Signal Processing*. Academic Press, pp. 125–151, (2025). http://doi.org/10.1016/B978-0-443-22158-3.00006-5

[11] X. Feng *et al.*, "Feasibility of fully automatic assessment of cervical canal stenosis using MRI via deep learning," *Quantitative Imaging in Medicine and Surgery*, vol. 15, no. 9, pp. 8457–8470, (2025). http://doi.org/10.21037/qims-2025-67

[12] C.-W. Chen, H.-H. Tsai, C.-Y. Yeh, C.-K. Yang, H.-K. Tsou, P.-Y. Leong, and J. C.-C. Wei, "Application of artificial intelligence in rheumatic disease classification: An example of ankylosing spondylitis severity inspection model," *Annals of Medicine*, vol. 57, no. 1, Art. no. 2512131, (2025). https://doi.org/10.1080/07853890.2025.2512131

[13] M. A. Larhmam, M. Benjelloun, and S. Mahmoudi, "Vertebra identification using template matching model and k-means clustering," *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, pp. 177–187, (2014). http://doi.org/10.1007/s11548-013-0927-2

[14] F. Muhammad et al., "Magnetization transfer MRI (MT-MRI) detects white matter damage beyond the primary site of compression in degenerative cervical myelopathy using a novel semi-automated analysis," *Computers in Biology and Medicine*, vol. 197, pt. B, Art. no. 111083, (2025). http://doi.org/10.1016/j.compbiomed.2025.111083

[15] K. Nispel, T. Lerchl, G. Gruber, H. Moeller, R. Graf, V. Senner, and J. S. Kirschke, "From MRI to FEM: an automated pipeline for biomechanical simulations of vertebrae and intervertebral discs," *Frontiers in Bioengineering and Biotechnology*, vol. 12, Art. no. 1485115, (2025). http://doi.org/10.3389/fbioe.2024.1485115

[16] E. Kondrateva *et al.*, "Negligible effect of brain MRI data preprocessing for tumor segmentation," *Biomedical Signal Processing and Control*, vol. 96, part A, p. 106599, Oct. (2024). http://doi.org/10.1016/j.bspc.2024.106599

[17] B. De Leener *et al.*, "SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data," *NeuroImage*, vol. 145, pp. 24–43, (2017). http://doi.org/10.1016/j.neuroimage.2016.10.009

[18] S. Patel and D. Kumar, "Predictive identification of oral cancer using AI and machine learning," *Oral Oncol. Rep.*, vol. 13, (2025). https://doi.org/10.1016/j.oor.2024.100697

[19] N. Bhagwat et al., "Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses," *GigaScience*, vol. 10, no. 1, (2021). https://doi.org/10.1093/gigascience/giaa155

[20] K. B. De Raad et al., "The effect of preprocessing on convolutional neural networks for medical image segmentation," *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, (2021). https://doi.org/10.1109/ISBI48211.2021.9433952

[21] M. Juneja, N. Aggarwal, S. K. Saini, S. Pathak, M. Kaur, and M. Jaiswal, "A comprehensive review on artificial intelligence-driven preprocessing, segmentation, and classification techniques for precision furcation analysis in radiographic images," *Multimedia Tools and Applications*, vol. 84, pp. 21467–21520, (2024) https://doi.org/10.1007/s11042-024-19920-3

[22] G. Ghobrial and C. Roth, "Deep learning-based automated segmentation and quantification of the dural sac cross-sectional area in lumbar spine MRI," *Frontiers in Radiology*, vol. 5, Art. no. 1503625, (2025). https://doi.org/10.3389/fradi.2025.1503625

[23] X. Chang, X. Cai, Y. Dan, Y. Song, Q. Lu, G. Yang, and S. Nie, "Self-supervised learning for multi-center magnetic resonance imaging harmonization without traveling phantoms," *Physics in Medicine & Biology*, vol. 67, no. 14, (2022). https://doi.org/10.1088/1361-6560/ac7b66

[24] M. Seif, T. Leutritz, S. Schading, T. Emmengger, A. Curt, N. Weiskopf, and P. Freund, "Reliability of multi-parameter mapping (MPM) in the cervical cord: A multi-center multi-vendor quantitative MRI study," *NeuroImage*, vol. 264, p. 119751, (2022). https://doi.org/10.1016/j.neuroimage.2022.119751

[25] W. Y. Zhang and H. J. Chen, "Applications and prospects of artificial intelligence in cervical spine surgery," *Zhonghua Wai Ke Za Zhi*, vol. 63, no. 5, pp. 383–388, (2025). http://doi.org/10.3760/cma.j.cn112139-20250104-00007

[26] T. W. Henry *et al.*, "Is the atlantoaxial level overlooked in the radiologic interpretation of cervical magnetic resonance imaging?" *Clin. Spine Surg.*, vol. 38, no. 10, pp. E568–E572, 2025). http://doi.org/10.1097/BSD.0000000000001805

[27] M. Pekmezci, A. A. Theologis, R. Dionisio, R. Mackersie, and R. T. McClellan, "Cervical spine clearance protocols in Level I, II, and III trauma centers in California," *Spine J.*, vol. 15, no. 3, pp. 398–404, (2015). http://doi.org/10.1016/j.spinee.2014.12.142

[28] P. Woznicki, F. C. Laqua, A. Al-Haj, T. Bley, and B. Baeßler, "Addressing challenges in radiomics research: systematic review and repository of open-access cancer imaging datasets," *Insights Imaging*, vol. 14, no. 1, p. 216, (2023). http://doi.org/10.1186/s13244-023-01556-w

[29] J. P. Harper, G. R. Lee, I. Pan, X. V. Nguyen, N. Quails, and L. M. Prevedello, "External validation of a winning artificial intelligence algorithm from the RSNA 2022 cervical spine fracture detection challenge," *AJNR Am. J. Neuroradiol.*, vol. 46, no. 9, pp. 1852–1858, (2025). http://doi.org/10.3174/ajnr.A8715

[30] C. J. Entenmann, K. Kersting, P. Vajkoczy, and A. Zdunczyk, "Closing the diagnostic gap: A narrative review of recent advances in functional MRI diagnostics in spinal cord injury," *Brain Spine*, vol. 5, p. 104283, (2025). http://doi.org/10.1016/j.bas.2025.104283

[31] M. V. C. Souza, R. Z. V. Costa, and K. E. P. Pinho, "Comparative analysis of parameters of magnetic resonance examinations for the accreditation process of a diagnostic image center in Curitiba/Brazil," in *XXVII Brazilian Congress on Biomedical Engineering*, Springer International Publishing, (2022). [Online]. Available: https://www.springerprofessional.de/comparative-analysis-of-parameters-of-magnetic-resonance-examina/20313580 (Accessed: Mar. 23, 2026)

[32] V. A. Magnotta, L. Friedman, and F BIRN, "Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study," *J. Digit. Imaging*, vol. 19, no. 2, pp. 140–147,(2006). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3045184

[33] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, (1990). https://ieeexplore.ieee.org/document/56205

[34] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recogn.*, vol. 38, no. 12, pp. 2270–2285, (2005). https://doi.org/10.1016/j.patcog.2005.01.012

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[36] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed., Pearson Education International, Pearson International Edition. (2008). [Online]. Available: https://sde.uoc.ac.in/sites/default/files/sde_videos/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods-ilovepdf-compressed.pdf. (Accessed: Mar. 23, 2026)

[37] J.-P. Fortin *et al.*, "Harmonization of multi-site diffusion tensor imaging data," *NeuroImage*, vol. 161, pp. 149–170, (2017). http://doi.org/10.1016/j.neuroimage.2017.08.047

[38] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging*, vol. 19, no. 2, pp. 143–150, (2000). http://doi.org/10.1109/42.836373

[39] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *arXiv e-prints*, (2017). http://doi.org/10.48550/arXiv.1703.10593

[40] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, (2009). http://doi.org/10.1016/j.ipm.2009.03.002

[41] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, (2009). http://doi.org/10.1109/MSP.2008.930649

[42] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, (1945). http://doi.org/10.2307/1932409

[43] K. J. Gorgolewski *et al.*, "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Sci. Data*, vol. 3, p. 160044, Jun. (2016). http://doi.org/10.1038/sdata.2016.44