All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems

Computer Methods and Programs in Biomedicine 250 (2024) 108200

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems

Silvia Seoni [a], Alen Shahini [a], Kristen M. Meiburger [a], Francesco Marzola [a], Giulia Rotunno [a], U. Rajendra Acharya [b,c], Filippo Molinari [a], Massimo Salvi [a,*]

[a] Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[b] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia
[c] Centre for Health Research, University of Southern Queensland, Australia

ARTICLE INFO

Keywords:
Multi-center studies
Multi-device studies
Image harmonization
Data preparation
Medical imaging
Artificial intelligence
Systematic review

ABSTRACT

Background and Objectives: Artificial intelligence (AI) models trained on multi-centric and multi-device studies can provide more robust insights and research findings compared to single-center studies. However, variability in acquisition protocols and equipment can introduce inconsistencies that hamper the effective pooling of multi-source datasets. This systematic review evaluates strategies for image harmonization, which standardizes appearances to enable reliable AI analysis of multi-source medical imaging.

Methods: A literature search using PRISMA guidelines was conducted to identify relevant papers published between 2013 and 2023 analyzing multi-centric and multi-device medical imaging studies that utilized image harmonization approaches.

Results: Common image harmonization techniques included grayscale normalization (improving classification accuracy by up to 24.42 %), resampling (increasing the percentage of robust radiomics features from 59.5 % to 89.25 %), and color normalization (enhancing AUC by up to 0.25 in external test sets). Initially, mathematical and statistical methods dominated, but machine and deep learning adoption has risen recently. Color imaging modalities like digital pathology and dermatology have remained prominent application areas, though harmonization efforts have expanded to diverse fields including radiology, nuclear medicine, and ultrasound imaging. In all the modalities covered by this review, image harmonization improved AI performance, with increasing of up to 24.42 % in classification accuracy and 47 % in segmentation Dice scores.

Conclusions: Continued progress in image harmonization represents a promising strategy for advancing healthcare by enabling large-scale, reliable analysis of integrated multi-source datasets using AI. Standardizing imaging data across clinical settings can help realize personalized, evidence-based care supported by data-driven technologies while mitigating biases associated with specific populations or acquisition protocols.

1. Introduction

The current era of big data is characterized by an unprecedented volume and variety of digital information, which has revolutionized and is continuing to revolutionize the way we collect, analyze, and derive insights from data to inform decision-making across diverse fields [1]. This is especially true in the field of medical imaging, where we can notice an increasing trend in combining data from multiple centers and acquisition systems for high-impact studies [2,3]. The integration of data from different sources in medical imaging is fundamental in that it allows the demonstration of generalizability and applicability of specific methods across diverse datasets [4–6]. Moreover, including a diverse and increased sample size can improve the statistical power of study results, allowing the detection of smaller but still clinically relevant effects. However, a significant challenge in such multi-centric and multi-device studies, defined as those involving several centers or devices equal to or higher than 2, is the presence of unwanted variability in the acquired images. This variability encompasses a range of factors,

* Corresponding author at: Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi, 24 10129 Turin, Italy.
E-mail address: massimo.salvi@polito.it (M. Salvi).

https://doi.org/10.1016/j.cmpb.2024.108200
Received 27 January 2024; Received in revised form 20 April 2024; Accepted 22 April 2024
Available online 23 April 2024

including differences in image intensity, variations in image codification and range (such as uint8, int16, etc.), disparities in pixel spacing, and other related aspects, such as any other relevant factors that could impact the consistency or comparability of the images. These variations can arise due to differences in imaging protocols, hardware specifications, and environmental conditions.

Minimizing unwanted variability in acquired images across multiple centers is crucial for ensuring reliable and consistent results, especially when employing artificial intelligence (AI) systems [7,8]. AI-based methods have shown increasingly powerful results for various tasks in medical image analysis, such as segmentation and classification. Still, they can present the drawback of over-learning on the dataset that is provided during the training phase, potentially creating biases, and making it struggle in the ability to generalize and provide satisfactory results on new datasets. Hence, the availability of diverse cases from multiple centers during the training and testing phases of AI-based methods is crucial as it represents the real-world clinical scenario. Therefore, while it is paramount to provide diverse cases from different sources, it is equally important to harmonize the image data to minimize the impact of unwanted variability and enable meaningful comparisons and analysis [9]. Hence, image harmonization aims to create a more consistent and standardized dataset for further analysis and evaluation, helping to ensure a fair and unbiased representation of the dataset and preventing perpetuating or amplifying existing biases. In this context, image harmonization refers to the specific techniques employed to standardize and harmonize the appearance and characteristics of images acquired from different sources or devices. It is an essential component of the broader data preparation process, which may also include data curation, pre-processing steps beyond harmonization, and other data management tasks.

Fig. 1 shows a typical pipeline commonly used in multi-centric or multi-device studies. In this pipeline where images are collected and subsequently undergo this fundamental harmonization process.

But how is image harmonization achieved in multi-centric and multi-device studies? Various methods exist, including protocol harmonization and image normalization. Protocol harmonization involves standardizing of acquisition protocols to obtain more consistent results [10]. Even if a strict protocol harmonization between centers exists, it may still be necessary to harmonize the images through image normalization techniques, often referred to as pre-processing, denoising, normalization, or standardization. These techniques focus on adjusting the acquired images to a common reference or standard, aiming to mitigate the differences in image appearance, intensity, and spatial characteristics. By applying these techniques, researchers can ensure that the images from different centers and devices have similar characteristics, facilitating meaningful comparisons and analysis [11].

1.1. Related reviews

Several previous reviews have explored image harmonization and pre-processing techniques in medical imaging, but they have certain limitations. Some focused exclusively on specific modalities like radiology [10,12,13] or digital pathology [9,14], while others covered a narrow scope of techniques, such as color normalization [14]. These reviews highlighted the importance of harmonization methods in improving model performance and generalizability, but they lacked a comprehensive, multi-modality perspective:

- Vasuki et al. [12] “A survey on image preprocessing techniques for diverse fields of medical imagery”: This review provides a survey of image preprocessing techniques across diverse fields of medical imagery, including radiology, nuclear medicine, and fundus imagery. However, it is outdated only covers 14 studies, and does not discuss AI models or the impact of preprocessing.
- Makandar et al. [13] “A Review on Preprocessing Techniques for Digital Mammography images”: This review focuses on preprocessing techniques in multi-centric studies within radiology. However, its scope is limited to mammography images.
- Mali et al. [10] “Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods”: This review summarizes image harmonization techniques in radiology. However, its focus is primarily on the impact of radiomics analysis.
- Pinto et al. [15] “Harmonization of Brain Diffusion MRI: Concepts and Methods”: This review concentrates explicitly on image harmonization in MR of the brain.

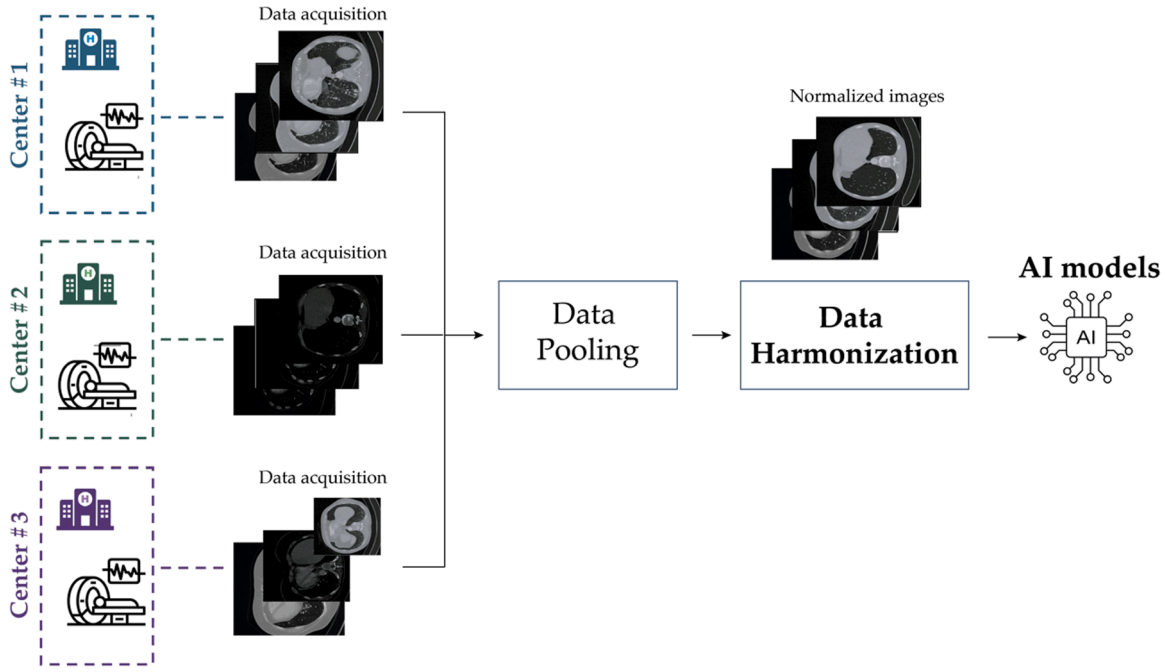

Fig. 1. Typical pipeline used in multi-centric or multi-device studies. Images are collected from different centers and pooled together for data harmonization to reduce variability. Then, normalized images are fed into the AI model for training or testing purposes.

- Salvi et al. [9] "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis" describes different types of pre-processing techniques limited to digital pathology.
- Tosta et al. [14] "Computational normalization of H&E-stained histological images: Progress, challenges and future potential": This review focuses explicitly on color normalization techniques in digital pathology.

This systematic review aims to provide a comprehensive overview of image harmonization approaches employed in multi-centric and multi-device studies within the healthcare domain. By analyzing studies published between 2013 and 2023, we aim to identify the most commonly used and effective techniques for harmonizing imaging data. Several image modalities are included and discussed, such as radiology imaging (Computed Tomography – CT, Magnetic Resonance Imaging – MRI, and mammography), nuclear imaging (Positron Emission Tomography – PET and Single Photon Emission Computed Tomography - SPECT), optical imaging (Optical Coherence Tomography – OCT, digital pathology, and fluorescence microscopy), ultrasound (US), and dermoscopy imaging. Additionally, we assess the impact of the various harmonization strategies on the outcomes and performance of AI models in multi-center studies.

1.2. Image harmonization approaches

Image harmonization approaches in medical imaging can vary greatly depending on the clinical application and available modalities. Fig. 2 provides an overview of common image harmonization methods. Broadly, most techniques fall into one of 5 categories:

- *Grayscale normalization*: grayscale normalization aims to standardize the intensity levels of grayscale images across different sources or imaging devices. It ensures consistent brightness and contrast, facilitating fair comparisons and analysis of image features.
- *Resampling:* these techniques involve scaling, resizing, or interpolation to harmonize images with different spatial resolutions. These methods enable alignment and consistency in size and spatial properties, improving compatibility and comparability between images.
- *Color normalization*: this normalization aims to standardize color appearance across images captured under different lighting conditions or using different color representations. It ensures consistent color characteristics, facilitating accurate analysis and interpretation of color-based features or lesions.
- *Denoising*: these methods aim to reduce noise or unwanted artifacts in images, enhancing their quality and improving the accuracy of subsequent analysis or interpretation. By removing noise, these techniques enhance image clarity and facilitate more reliable feature extraction or detection.
- *Contrast enhancement*: these enhancement techniques aim to adjust the contrast levels of an image to improve the visibility and differentiation of objects or structures. By enhancing the contrast, these techniques help reveal finer details and improve the interpretability of image features.

These image harmonization techniques play a crucial role in the landscape of multi-centric or multi-device studies, ensuring consistency, comparability, and enhanced image quality. They are fundamental for AI methods as they contribute to the reliability, accuracy, and generalizability of the developed models. By harmonizing the image data in input, these techniques facilitate accurate and reliable analysis, interpretation, and diagnosis, allowing the AI model to effectively focus on learning crucial image aspects that may differentiate between a healthy and a pathological subject, and hence ultimately enhancing the effectiveness of healthcare applications and research.

1.3. Image modalities

The review encompasses a range of image modalities in healthcare. The following list provides an overview of the included modalities along with their descriptions:

- Radiology: this modality includes Computed Tomography (CT) imaging, Magnetic Resonance (MR) imaging, and mammography [16, 17]. As research in CT imaging grows, the demand for multi-center or multi-device studies has arisen. However, these studies face challenges due to differences in image acquisition parameters and protocols among various imaging centers or devices. Variations in factors like tube current, voltage, slice thickness, and reconstruction algorithms can affect image quality and introduce data inconsistencies in CT imaging. MR imaging also often struggles with a lack of uniformity in its acquisition protocols. This variability poses significant challenges in ensuring the consistency of MR imaging data across different centers and studies, potentially affecting the accuracy and reliability of subsequent analyses, particularly in radiomics where quantitative feature extraction is vital. Furthermore, standardizing acquisition and reconstruction protocols is primarily feasible only for prospectively collected data. Mammography,

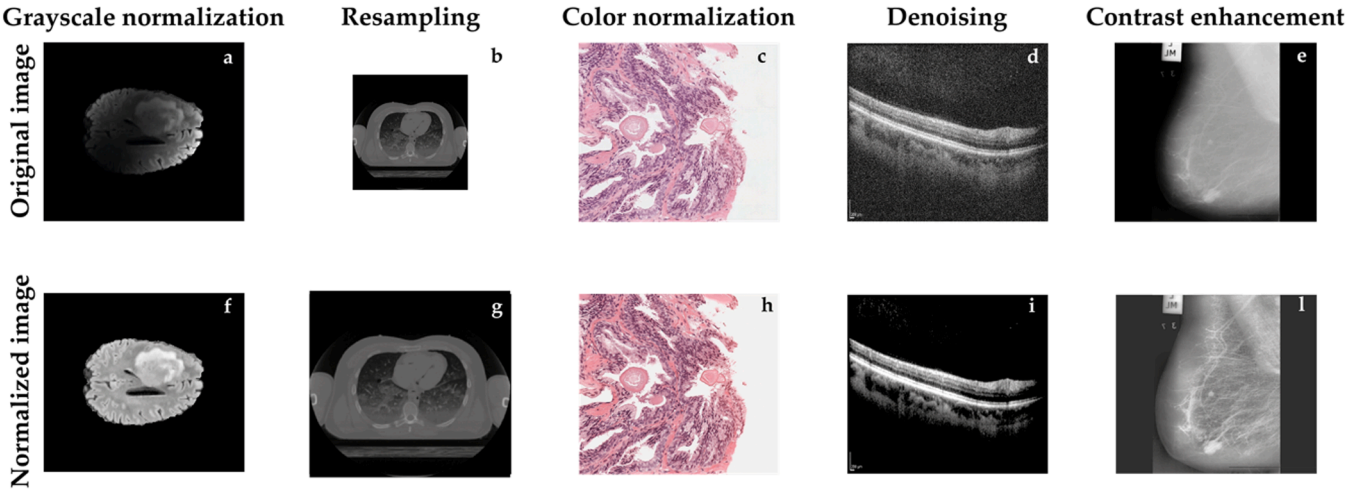

Fig. 2. Example data preparation approaches in healthcare: (a, f) grayscale normalization in MR images, (b, g) resampling in CT images, (c, h) color normalization in digital pathology, (d, i) denoising in OCT images, (e, l) contrast enhancement in mammographic images.

a critical method for breast examination and cancer screening, utilizes low-dose X-rays to detect early signs of breast cancer. However, it faces challenges such as high rates of false positives, leading to unnecessary biopsies, and false negatives, resulting in missed diagnoses. In multi-center and multi-device studies, these challenges are exacerbated by differences in image acquisition and processing across clinical settings. Variability in protocols, equipment, calibration, and techniques can hinder result comparability and reliability.

- Nuclear imaging: this modality involves Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) [18]. Using scanner models and acquisition protocols across sites poses significant challenges for image harmonization and pooling. Factors such as differences in scanner calibration, reconstruction algorithms, attenuation, and scatter correction methods can introduce systematic biases between data acquired on different scanners. Patient movement, variation in radiotracer dose, and uptake time also impact quantitative values derived from PET and SPECT images. Without proper harmonization, combining data from multiple centers risks violating assumptions of test equivalence and comparisons across treatment groups. This hampers the pooling of imaging cohorts for large-scale analysis.
- Optical imaging: this modality encompasses digital pathology, fluorescence imaging, optical coherence tomography (OCT), and OCT angiography (OCTA). Digital pathology involves scanning and digitizing histological slides for computer-based analysis, enabling detailed examination of tissue samples [19]. The need for multi-centric studies in digital pathology arises from the desire to validate findings across different institutions and ensure the generalizability of results. However, multi-centric studies pose challenges due to the inherent variability in data acquisition, staining techniques, and imaging protocols. Fluorescence imaging is crucial for studying diseases, discovering drugs, and personalizing medicine. Multi-center studies using this method are growing, but they struggle with image consistency due to variations in devices, protocols, and analysis methods. These differences make it challenging to compare results accurately, potentially leading to inconsistent conclusions. OCT/OCTA captures high-resolution cross-sectional images of biological tissues, providing valuable insights into morphological structures, such as retinal layers (OCT) [20] and blood vessels (OCTA) [21]. OCT and OCTA are essential imaging modalities especially for ophthalmology and dermatological applications [22, 23]. However, performing extensive multi-centric clinical studies with OCT/OCTA can be challenging due to variability in acquisition settings and equipment between sites. Different OCT systems have a range of resolutions, wavelength sources, and scanning protocols that can impact imaging quality and lead to inconsistencies in measurements and diagnoses [24].
- Ultrasound (US): ultrasound imaging is widely used for diagnostic purposes. It is particularly valuable in obstetrics, cardiology, and musculoskeletal imaging [25–27] However, ultrasound images can vary significantly depending on the acquisition settings, operator, and US device used. These sources of variability present challenges for multi-centric studies that aim to pool ultrasound data from multiple clinical sites. Factors like scanner brand, transducer model, imaging frequencies, focal zones, and acquisition depth can all impact the resolution and appearance of ultrasound images. Additionally, differences in how operators position the transducer and adjust gain settings contribute to variability.
- Dermoscopy: This technique, used in dermatology, involves imaging for skin condition diagnosis. Dermoscopic images can be taken with smartphones or digital cameras, but in clinics, dermatoscopes are usually used to compress lesions and capture epiluminescence images [28,29]. The need for multi-centric studies in dermatology arises from the desire to gather a broader and more diverse dataset to enhance research findings and improve patient care. However, conducting studies across multiple centers introduces challenges related to data variability. These challenges primarily stem from variations in the acquisition settings, imaging devices, resolution, and lighting conditions used in different centers.

This review explores a wide range of analyses used in multi-centric and multi-device studies, focusing on the techniques that can be used for image harmonization. Commonly used image harmonization techniques, such as those listed in Section 1.2, can be implemented in several ways, and are not limited to one simple algorithm or implementation. For clarity purposes, we separate the types of image harmonization methods into three macro-areas: math- or statistics-based, machine learning (ML) -based, and deep learning (DL) -based approaches. In this review, each image modality is analyzed, and the types of image harmonization methods are divided into these three macro-areas. Furthermore, the clinical tasks that are confronted in the analyzed studies may differ, including classification, detection, segmentation, prediction, and image quality assessment. Classification involves assigning images to predefined classes or categories, while detection focuses on identifying specific objects or features within the images. Segmentation aims to delineate and separate different regions or structures of interest in the images. Prediction involves estimating or forecasting certain properties or outcomes based on the harmonized image data. Additionally, image quality assessment plays a crucial role in evaluating the fidelity and reliability of harmonized images.

In the following sections, we will explore the various techniques and tasks involved in multi-centric and multi-device studies for different imaging modalities. We will specifically focus on the importance and impact of image harmonization for AI applications in the analyzed studies.

2. Methods

To select the most relevant articles, we closely followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

2.1. Literature search strategy

This review focuses on articles published between 2013 and 2023 investigating the use of machine and deep learning methods for medical image analysis. The decision to analyze works from the past decade ensures a focus on recent advancements and trends in the field, providing a comprehensive examination of developments while capturing recent innovations and methodologies. A literature search was conducted in October 2023 across scientific databases including Scopus, the Institute of Electrical and Electronics Engineers (IEEE), and PubMed. The search strategy employed a Boolean approach, combining various keywords such as “Multi-centric”, “Machine learning”, “Deep learning”, “Detection”, “Classification”, “Segmentation”, “Prediction”, “Diagnosis”, “Healthcare”, with “CT”, “PET”, “MRI”, “US”, “Photoacoustic”, “Digital Pathology”, “Ultrasound”, “Dermoscopy”, “OCT”, “Fluorescence” in different combinations.

This initial search returned 262 articles. Articles were then screened to remove duplicates ($n = 11$), as well as books, abstracts, and conference proceedings. The remaining studies were further assessed based on journal quality, focusing on those published in top-quartile (Q1) journals according to impact factor metrics. The assessment of the remaining studies was then based on the following criteria:

(i) A description of multi-centric studies for image classification, detection, or segmentation should be included.
(ii) A description of methods based on machine learning or deep learning models should be included
(iii) Written in English.

Articles that did not meet these criteria were excluded, and the pilot

studies, works published before 2013, or articles not available in full text. This review process resulted in a final set of 100 studies focusing on the application of multi-centric and/or multi-device strategies in healthcare. To ensure the reliability of our findings, we conducted a risk of bias assessment with two independent reviewers. One reviewer performed the initial literature search, while the other assessed studies for inclusion, thereby mitigating potential bias in the conclusions. Fig. 3 depicts the utilization of the PRISMA guideline for the systematic article screening and selection process.

3. Results

3.1. Radiology imaging

3.1.1. MR imaging

Table A1 summarizes the studies discussed in this section and the effects of the implemented strategies. As shown in Fig. 4, given the physical meaning of voxel intensity in MR, almost all the methods presented in this section are related to grayscale normalization, and only one method is primarily related to denoising techniques. Only one method primarily focuses on denoising. The most commonly used techniques for grayscale normalization are based on mathematical methods, although machine learning and deep learning approaches are also utilized. Furthermore, the majority of the downstream tasks are related to the segmentation of anatomical structures, but classification and evaluation of image quality are also represented. Grayscale normalization enhanced the comparability of MR images acquired from different centers or with different imaging protocols by addressing differences in voxel intensities across scanners and protocols. This highlights the benefit of grayscale normalization for improving the performance of algorithms applied to multi-center or multi-protocol MRI data.

Several methods have been developed to address standardization challenges, mainly focusing on grayscale normalization. Carrè et al. [30] investigated the impact of three intensity normalization methods (Nyul, WhiteStripe, z-score) combined with two discretization techniques. They demonstrated that intensity normalization enhanced the robustness of first-order radiomics features, with mean balanced accuracy for tumor grade classification increasing from 67 % to as high as 82 %. Ji et al. [31] developed a cross-vendor bi-parametric radiomic model for differentiating between benign and malignant prostate lesions employing a combination of T2-Weighted Imaging and Apparent Diffusion Coefficient measures. They applied z-score normalization achieving an AUC of 0.93 with the inner test set and 0.88 in the outer test. Alnowami et al. [32] employed a DenseNet for classifying brain tumors analyzing approximately 4314 MRI images across four classes (normal, glioma, meningioma, and pituitary tumor). Their research highlighted the effectiveness of intensity normalization techniques, such as WhiteStripe and z-Score, improving average classification accuracy from 72.1 % up to 96.52 %. Foltyn-Dumitru et al. [33] focused on the impact of N4 bias field correction on the generalizability of radiomic-based predictions for molecular glioma subtypes, using N4 followed by WhiteStripe (N4/WS) and z-score normalization (N4/z-score). Both N4/WS and N4/z-score significantly outperformed the other methods, achieving macro-average AUC scores ranging from 0.85 to 0.87 in external test sets, compared to 0.19 to 0.52 for the naive and N4 methods alone. Sun et al. [34] developed a histogram normalization method, comprising intensity scaling between low- and high-intensity regions. Through experimental validation in image segmentation, this method increased the Dice score up to 2.3 % compared to the unprocessed image. Pereira et al. [35] developed an automatic brain tumor segmentation method based on CNNs. Their image harmonization step involved filtering with N4ITK Bias Field filter and Nyul normalization achieving a mean Dice score of 84 % compared to 78 % using z-score normalization. Ou et al. [36] tackled the challenges of multi-site brain MRI analysis, mainly focusing on the variability in fields of view (FOVs) across different scanning sites and protocols. Their study introduced an atlas-based approach to FOV standardization improving Dice scores in downstream segmentation up to 25 %. Jacobsen et al. [37] applied four different intensity normalization methods during the pre-processing of a CNN-based method. They showed that histogram equalization methods outperformed unit distribution methods when evaluated using two external datasets with a median Dice improving from 85 % to 0.90 %. Modanwal et al. [38] proposed a novel normalization approach for breast MR images using a modified CycleGAN matching the desired intensity across two scanners and achieving a Dice score of 98 %, representing an 8 % increase over the baseline. Delisle et al. [39] introduced an adversarial and task-driven approach with a realism constraint to produce harmonized and realistic images across multiple datasets while optimizing for segmentation accuracy. They improved the mean Dice score by 5.6 % compared to the traditional min-max scaling. Koble et al. [40] investigated the efficacy of different histogram normalization techniques for segmenting of multispectral brain MR data. Their findings suggest that a properly adjusted Nyúl algorithm can produce a 0.5 % improvement in accuracy than a fine-tuned linear transform in DL-based segmentation. Albert et al. [41] tested six normalization techniques on multiple deep learning tasks. They suggest normalization in neural networks aids by incorporating prior knowledge and is more impactful on small, inhomogeneous datasets. It significantly influences classification and regression tasks over-segmentation. In single-center data training, external evaluation showed no significant difference in Dice scores. Reiche et al. [42] developed a framework for multi-institutional FLAIR MR datasets, focusing on preserving the appearance of white matter lesions (WML) while normalizing intensity. Their approach involved denoising, background subtraction, bias field correction, and a novel histogram-based intensity standardization. They improved the KL divergence between the dataset from 1.013 to 0.094. This work was then evaluated by Ghazvanchahi et al. [43] who investigated intensity standardization methods for WML using DL-based segmentation in multi-centric FLAIR MRI. They assessed various normalization techniques, including IAMLAB [42] and proposed an

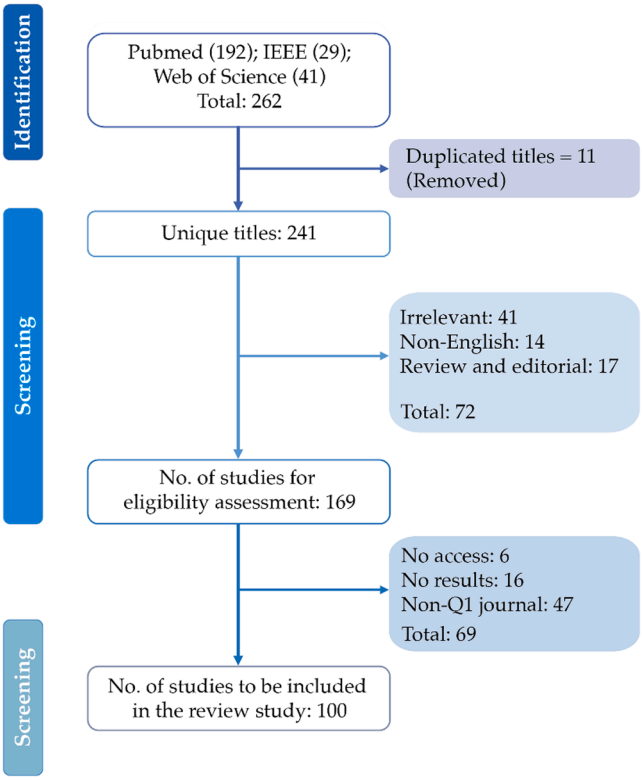

Fig. 3. Selection of relevant articles based on PRISMA guidelines.

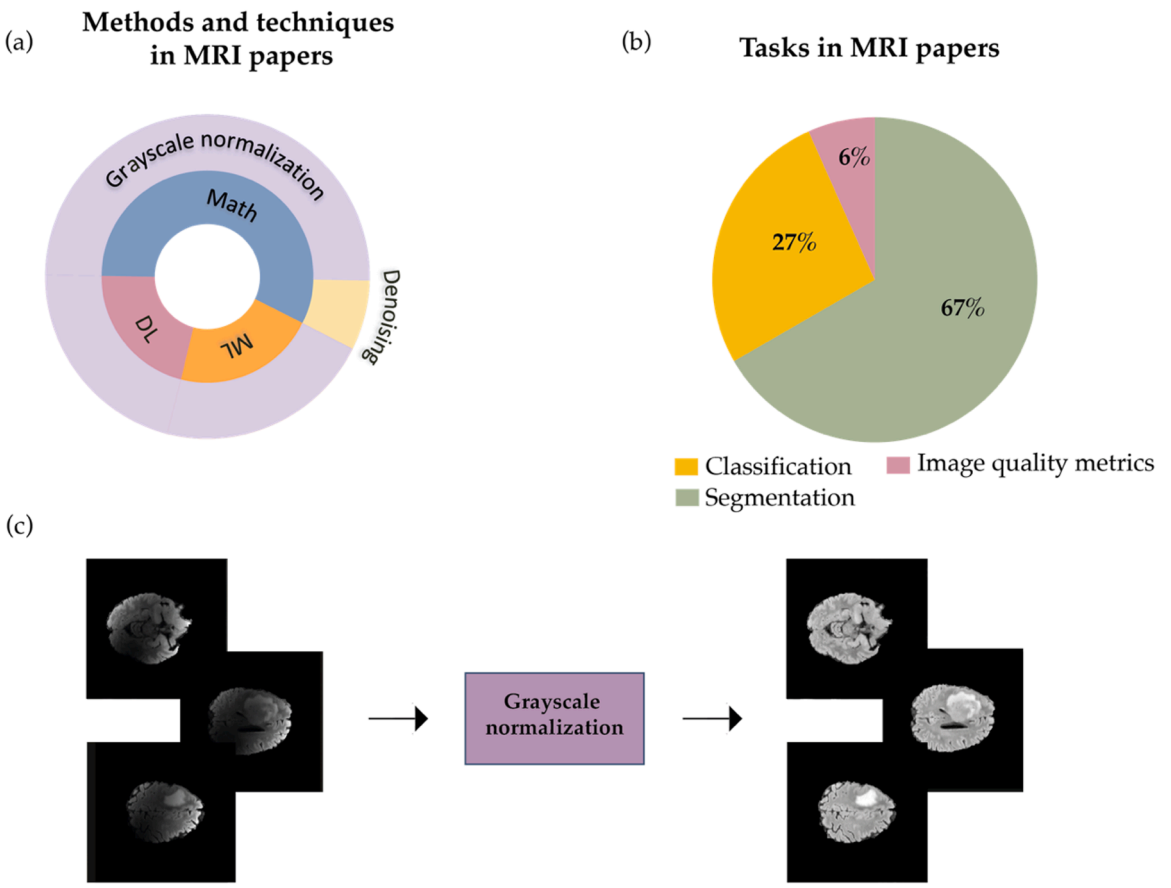

Fig. 4. Summary of the studies ($n = 14$) on image harmonization in MR. (a) Techniques employed in MR imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric MR imaging studies. (c) Example of image harmonization in MR imaging. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

ensemble model combining predictions from these methods. They observed significant improvements in sensitivity, resulting in 69 % for IAMLAB and 78 % for the ensemble method compared to 66 % using original data.

In MR imaging, image harmonization techniques have shown promising results in enhancing the robustness of radiomics features and improving the performance of downstream tasks. Grayscale normalization methods, such as z-score, have been widely employed, with studies reporting significant improvements in classification accuracy, ranging from 67 % to 96.52 % [30,32]. Combining of these techniques with other preprocessing steps, such as N4 bias field correction, has further enhanced the generalizability of radiomic-based predictions, with macro-average AUC scores reaching 0.85 to 0.87 in external test sets [33]. Cross-vendor models and atlas-based approaches have also contributed to the standardization of MR imaging data across different scanning sites and protocols, with FOV standardization improving Dice scores in downstream segmentation by up to 25 % [36]. Histogram normalization and bias field correction have demonstrated their effectiveness in image harmonization, with studies reporting improvements in Dice scores ranging from 2.3 % to 8 % [34,38]. These advancements contribute to the broader goal of facilitating multi-centric and retrospective studies in radiomics research, enhancing our understanding of patients' diseases.

3.1.2. CT imaging

Table A2 summarizes the studies discussed in this section, along with the effects of the implemented strategies. Grayscale normalization refers to the harmonization of voxel values depending on the statistical distribution of the intensity between centers and devices, while resampling techniques are linked to scaling methods. Fig. 5 shows the distribution of the works in this section. Grayscale normalization is typically performed using model-based techniques while statistical methods are used for resampling. In terms of downstream tasks, CT imaging wide and vast scope of tasks, given its wide applicability in clinics.

In the context of CT imaging, grayscale normalization is a crucial preprocessing step aimed at standardizing the intensity of voxels in retrospective studies. Li et al. [44] employed a generative model to adapt the images from 3 different devices A, B, and C to a target device T. Using the Wilcoxon-sum test, they computed the percentage of image features that were consistent between the different devices, which increased from 10.4 %, 18.2 % and 50.1 % for the unnormalized data to 93.5 %, 89.6 % and 77.9 % after normalization.

Resampling techniques apply statistical harmonization of the sampling and acquisition parameters of the images to reduce the differences between centers. Ligero et al. [45] resampled all acquisitions from two different centers to isometric voxels of $1 \times 1 \times 1$ mm^3 interpolating with splines and nearest neighbour methods. They applied ComBat [46] to perform batch correction improving K-means-based tumor type classification with respect to initial data (radiomics classification accuracy increased from 65.9 % to 73.2 %). Park et al. [47] improved the AUC of a Random Forest model for recurrence prediction of non-small cell lung cancer (NSCLC) from 0.70 to 0.80 using reconstruction kernels. They also standardized the voxel dimensions to 1 mm isovoxels through cubic interpolation. Finally, Tonneau et al.[48] resampled original voxels to 1-mm isometric voxels and applied a Laplacian of Gaussian filter to normalize the extraction of deep radiomics features. Through a generalization optimizing search framework, the survival rate prediction in NSCLC cancer improved the AUC from 0.52 to 0.63 in a validation

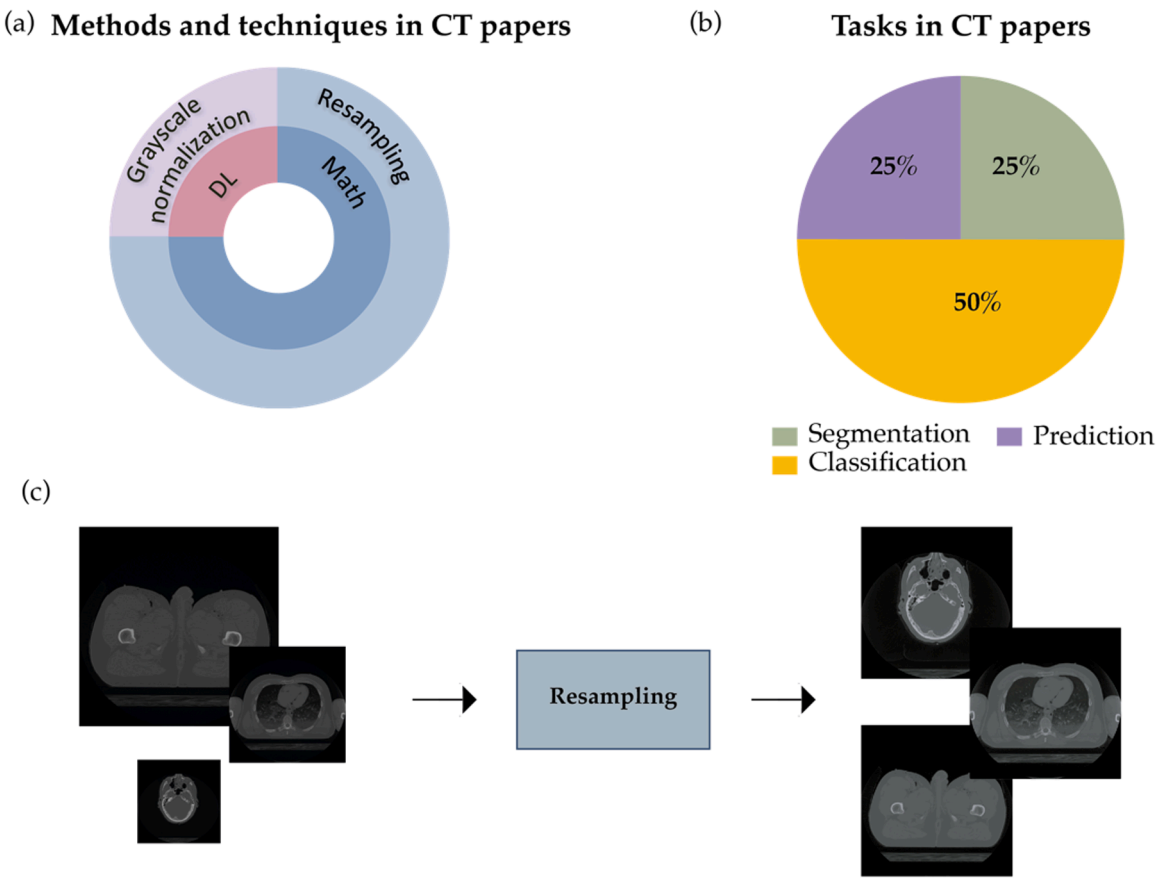

Fig. 5. Summary of the studies ($n = 4$) on image harmonization in CT. (a) Techniques employed in CT imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric CT imaging studies. (c) Example of image harmonization in CT imaging. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

cohort.

With the increasing prominence of multi-centric or multi-device studies in CT imaging, the challenges posed by the variability in image acquisition parameters and protocols have become more apparent. The normalization of CT images in multi-centric studies not only enhances the consistency and comparability of data across different centers but also plays a crucial role in improving the accuracy and reliability of diagnostic and predictive models. Grayscale normalization techniques have demonstrated significant improvements in feature consistency across different devices, with the percentage of consistent image features increasing from as low as 10.4 % to as high as 93.5 % after normalization [44]. Resampling techniques, which involve statistical harmonization of sampling and acquisition parameters, have also shown promise in reducing inter-center differences. The standardization of voxel dimensions has led to improvements in tumor type classification accuracy, increasing from 65.9 % to 73.2 % [46]. The combination of resampling and filtering techniques has also been shown to enhance the performance of deep radiomics features in survival rate prediction for NSCLC, with the AUC improving from 0.52 to 0.63 in a validation cohort [48]. As radiomics advances, implementing effective image harmonization techniques will be crucial for fully leveraging the information contained within the CT images.

3.1.3. *Mammography*

Table A3 summarizes the studies discussed in this section, along with the effects of the implemented strategies. Fig. 6 illustrates the distribution of techniques used and how they lead to different downstream evaluation tasks. The classification was the most common task examined, along with image quality metrics, which aligns with the clinical application of mammography in cancer screening. Contrast enhancement techniques based on statistical methods primarily comprise the landscape of harmonization approaches in mammography. More recently, researchers have also explored noise reduction and grayscale normalization techniques. Nearly all harmonization methods are based on mathematical and statistical methods, except for one instance of machine learning for contrast enhancement and one instance of deep learning for denoising. The figure shows an example of how contrast-limited adaptive histogram equalization (CLAHE) can enhance image contrast in a multi-center dataset, demonstrating the potential for contrast enhancement techniques to improve harmonization across sites.

Contrast enhancement is vital role in improving the visibility of subtle abnormalities and enhancing diagnostic accuracy. In the study by Deng et al. [49], a novel mammogram enhancement algorithm (MIFS) is presented, which employs intuitionistic fuzzy sets to highlight fine details in mammograms more effectively achieving an average contrast value of 0.581 compared to 0.436 of the original data. Perez et al. [50] analyze a preprocessing pipeline on an exceptionally representative dataset obtained from 11 centers. The pipeline includes several steps, normalization of pixel values, histogram shifting, and linear stretching based on percentile values, showing an average increase in Dice score compared to unprocessed data of 23.5 %. Cao et al. [51] introduce the Breast Mass Detection Network (BMassDNet). This novel framework, enhanced with a truncation normalization method and adaptive histogram equalization for contrast improvement, shows true positive rates of 0.930 and 0.943 on the INbreast and DDSM datasets respectively, outperforming several methods.

More recently, DL techniques have been applied to the challenging

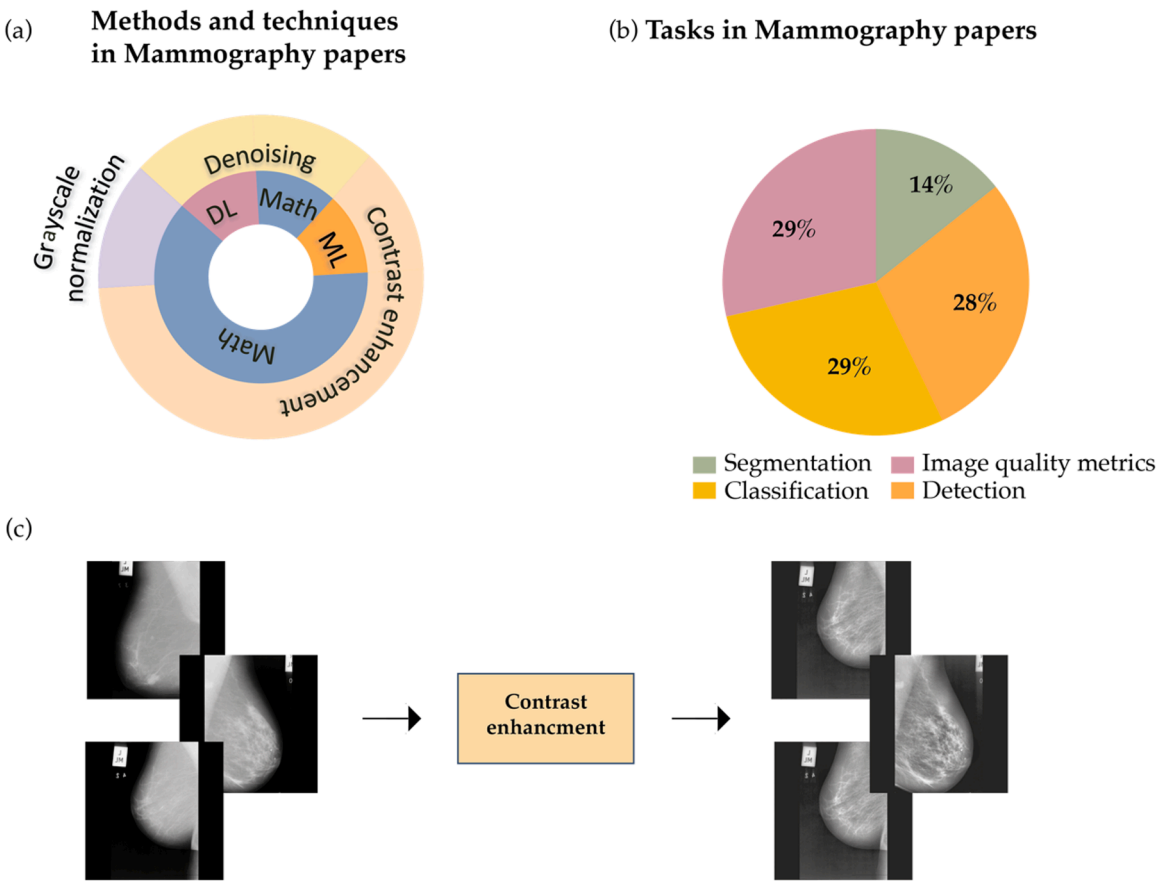

Fig. 6. Summary of the studies ($n = 4$) on image harmonization in mammography. (a) Techniques employed in mammography categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric mammography studies. (c) Example of image harmonization in mammography. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

task of harmonizing mammography images from different medical centers. Mechria et al. [52] explore the impact of denoising using a deep convolutional neural network, the DnCNN. They compare the classification performance with different denoising methods and with the original data, with improvements of 3.47 % overall accuracy, 5.34 % in specificity, and 0.56 % in sensitivity. Perre et al. [53] evaluated the impact of six different normalization methods on the performance of two CNNs, in the classification of mammographic images. They found that the effect of image normalization on the performance of the CNNs depends on which network is chosen, and that z-score normalization had the most positive impact, improving AUC from 0.763 to 0.786.

Effective image harmonization strategies, such as contrast enhancement techniques, denoising algorithms, and grayscale normalization, have improved the visibility of subtle abnormalities, diagnostic accuracy, and segmentation performance. Novel contrast enhancement algorithms have shown promising results, with an average contrast value of 0.581 compared to 0.436 in the original data [49]. Preprocessing pipelines incorporating normalization, histogram shifting, and linear stretching have led to an average increase in Dice score of 23.5 % compared to unprocessed data [50]. The impact of normalization methods on CNN performance has been evaluated, with z-score normalization demonstrating the most positive impact, improving AUC from 0.763 to 0.786 [53]. While traditional mathematical methods remain predominant due to their simplicity and direct applicability, the complexity of mammographic images, particularly in multi-centric and multi-device studies, has necessitated the exploration of more sophisticated approaches, such as ML and DL techniques. These advancements in image harmonization enhance the accuracy of breast cancer screening and improve the comparability and reliability of multi-centric and multi-device studies, leading to better patient outcomes.

3.2. Nuclear imaging

3.2.1. PET/SPECT imaging

Table A4 summarizes the studies that apply image harmonization in PET/SPECT while Fig. 7 shows the distribution of methods and validation tasks. Grayscale normalization is clearly the most important aspect for PET/SPECT given the technique's lower requirements for spatial resolution compared to the need for standardized pixel value interpretation, as pixel values relate to physical properties. Mathematical and statistical methods are predominantly utilized for grayscale normalization of PET/SPECT, with just one instance of a deep learning approach. The most common downstream validation tasks involve segmentation and classification. Fig. 7 also illustrates the benefits of grayscale normalization, showing improved harmonization in a sample PET dataset. The literature demonstrates that grayscale normalization techniques are crucial for PET/SPECT to standardize pixel-level intensities across acquisition devices and protocols.

One of the first issues to tackle to enable reliable voxel-wise statistical analysis and predefined-VOI–based automated anatomic labeling with nuclear imaging is the Spatial Normalization (SN) procedure. SN is a process that adjusts individual images to fit a standard template, addressing differences in size and shape among subjects often using an MR acquisition to align with the morphological content. Kang et al. [54] developed a deep learning–based SN method for amyloid PET imaging quantification that does not require MR or CT images. This approach was evaluated against the gold standard FreeSurfer [55], across three different amyloid PET radiotracers. In terms of correlation, the

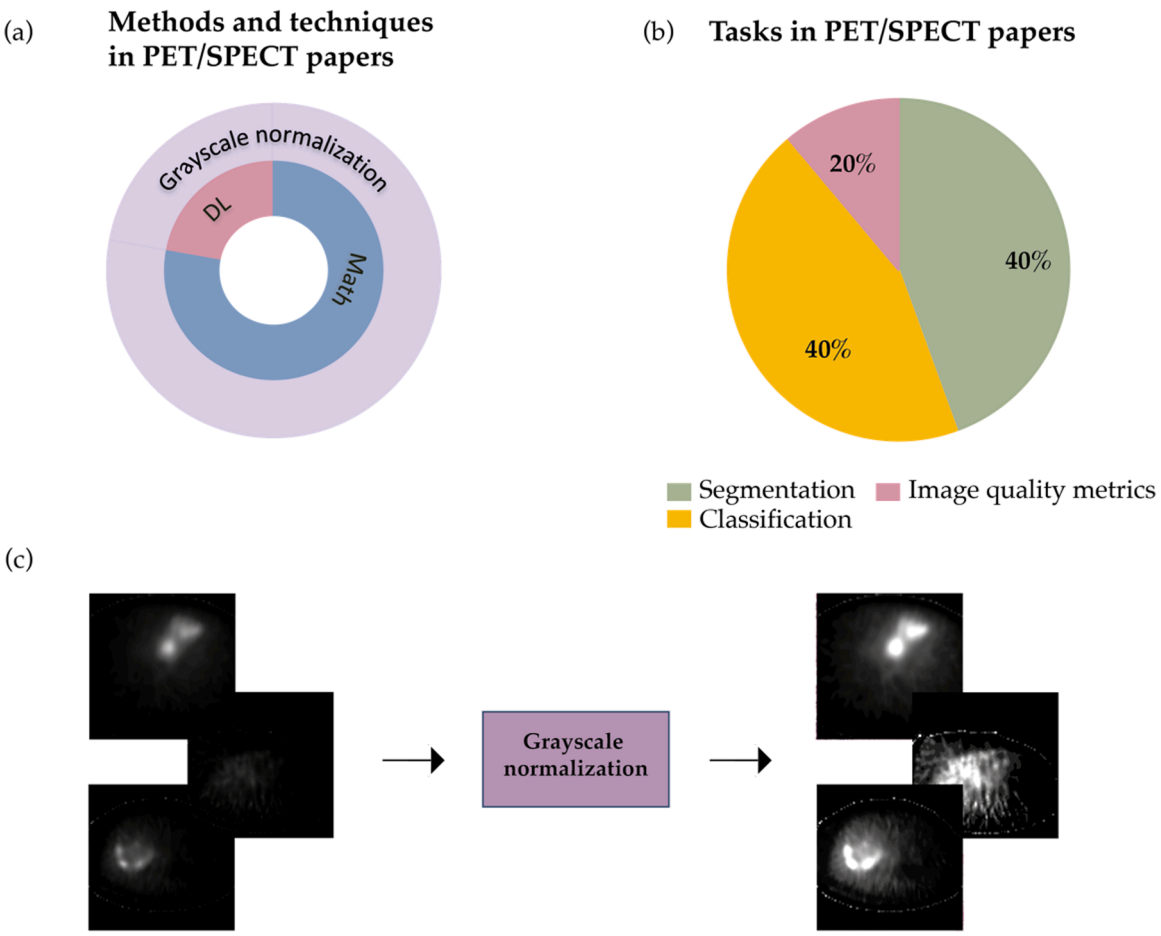

Fig. 7. Summary of the studies ($n = 4$) on image harmonization in PET/SPECT imaging. (a) Techniques employed in PET/SPECT imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric PET/SPECT imaging studies. (c) Example of image harmonization in PET/SPECT imaging. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

DNN-based PET SN method outperformed MRI-based PET SN with an R2 of 0.946 compared to 0.869 for the MR-based method. Similar steps of SN are typically found in downstream tasks of classification and segmentation of PET images. In Thiele et al. [56], PET brain images have been used in the context of a voxel-based classification system of neurodegenerative dementias. Images were spatially normalized to a template brain image using b-splines resulting in 91 × 109 × 91 isotropic 2 mm voxels, which were then smoothed with an isotropic Gaussian of 10 mm FWHM and normalized to a common median. To moderate inter-scanner variability, voxel-by-voxel scaling was applied based on "ratio images" calculated on controls. Classification accuracy using preprocessed data in the cross-scanner scenario improved from 79 % to 85 %. In their study, Lee et al. [57] evaluated the performance of CNNs in classifying Florbetaben amyloid brain PET scans, crucial for Alzheimer's disease diagnosis. Preprocessing steps such as spatial normalization, count normalization, and skull stripping were applied to both the internal and external datasets. The VGG 3D model achieved the highest performance with an AUC of 0.945 on an external dataset. Ren et al. [58] introduced two novel PET normalization methods, PET-Clip and PET-Sine, to enhance segmentation of head and neck tumors. PET-Clip clips the Standardized Uptake Value (SUV) values to a range of 0–5 and PET-Sine employs a sine transformation. Both methods aim to mitigate the impact of intensity variations across different PET scans. Using an ensemble of these normalization methods, they achieved a Dice score of 78 %, surpassing the baseline performance of 76 %.

The impact of image harmonization extends beyond improving the accuracy and reliability of quantitative measurements in PET/SPECT imaging. It also paves the way for advanced analysis techniques, such as voxel-wise statistical analysis, automated anatomic labeling, and classification systems for neurodegenerative diseases. In downstream tasks such as classification and segmentation, preprocessing steps often include spatial normalization, to address differences in size and shape among subjects. In a study on the classification of neurodegenerative dementias, preprocessing improved classification accuracy from 79 % to 85 % in a cross-scanner scenario [56]. Similarly, in classifying brain PET scans, a VGG 3D model achieved an AUC of 0.945 on an external dataset after applying preprocessing steps [57]. Through techniques like spatial normalization, deep learning-based methods, and novel normalization approaches, image harmonization ensures standardized interpretation of pixel values. It enhances the comparability of imaging data across different scanners and sites.

3.3. Optical imaging

3.3.1. Digital pathology imaging

Table A5 summarizes the works discussed in this section, along with the impact of the strategies employed. Techniques such as color normalization have been adopted to enhance the robustness of AI models in digital pathology. Color normalization, in brief, refers to the process of standardizing the color appearance of images, thereby reducing the effects of staining variations in the specific field of digital pathology, and is often referred to as stain normalization. It has been demonstrated that stain normalization not only improves the performance of AI algorithms [9] but also enhances the diagnostic accuracy of pathologists themselves [59]. Fig. 8 illustrates the distribution of the reported works in this section. Image harmonization techniques in

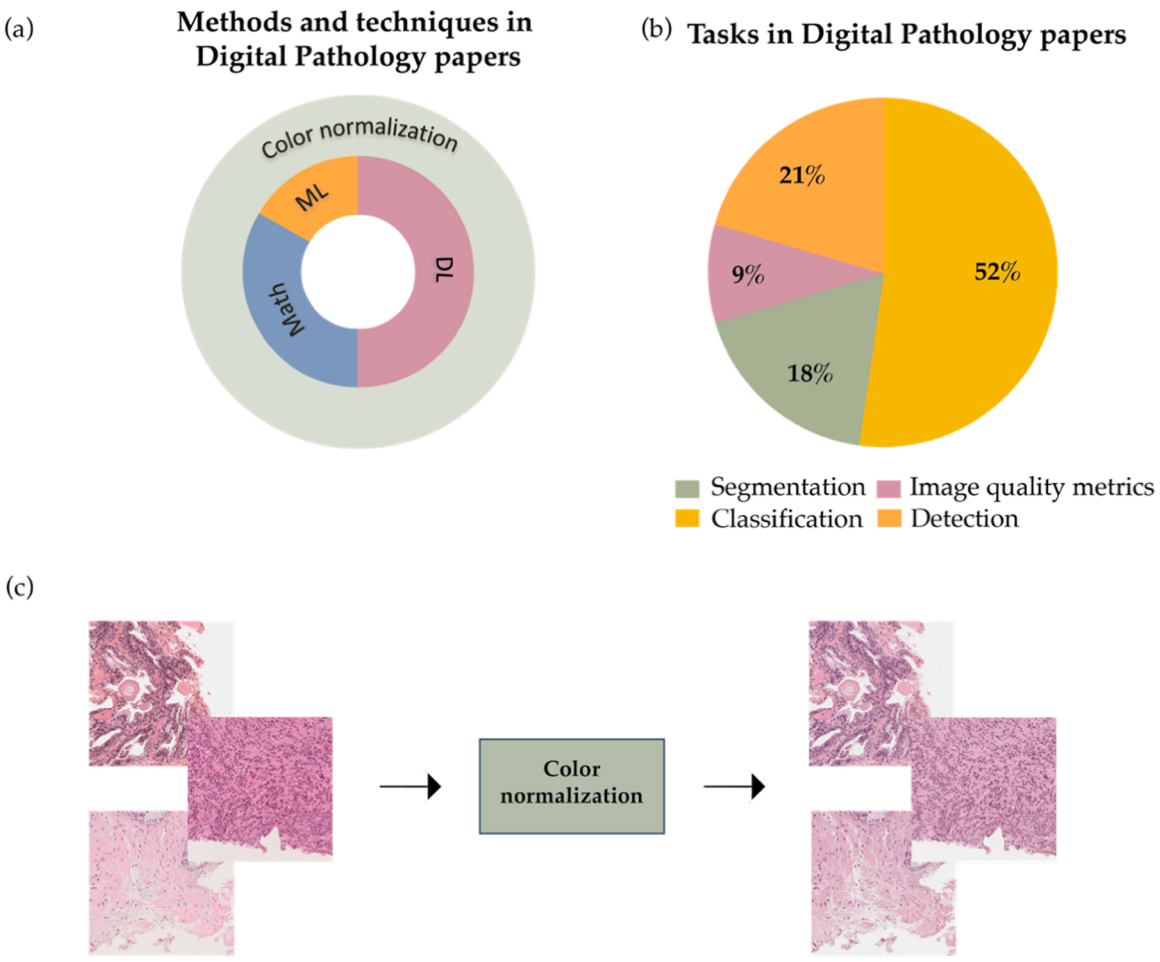

Fig. 8. Summary of the studies ($n = 38$) on image harmonization in digital pathology. (a) Techniques employed in digital pathology categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric digital pathology studies. (c) Example of image harmonization in digital pathology. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

digital pathology rely predominantly on color normalization, which is achieved mainly through DL methods, followed by statistical/mathematical methods, and finally, ML methods. Among the various tasks in digital pathology, classification tasks constitute the majority (54 % of the works), followed by segmentation (21 %), detection (18 %), and studies on image quality metrics. An example of color normalization in images from a multicentric dataset of digital pathology is shown in Fig. 8.

Initially, stain normalization approaches relied on mathematical and statistical techniques, such as color deconvolution, which separates the stain components in an image [60,61]. Color deconvolution is a method that separates the different stain components present in an image by exploiting the optical properties of stains and their interactions with light. It allows for estimating of the stain concentrations, which can then be used to normalize the color appearance of histopathology images [60]. Tam et al. [62] presented a histogram-based stain normalization method to enhance feature extraction in quantitative pathology. They achieved a 13 % increase in classification accuracy compared to unprocessed images. Anghel et al. [63] introduced a real-time stain normalization system using Macenko's stain vector estimation method, optimized for high-resolution whole-slide images. Their system delivers substantial speed enhancements compared to standard implementations and boosts classification accuracy by 5 %, even with low-quality input images. Salvi et al. [64] presented SCAN, a novel algorithm for stain separation and normalization of H&E slides. SCAN used cellular structure segmentation and clustering to estimate stain vectors and performed pixel-wise normalization. It outperformed other qualitative and quantitative methods both qualitatively and quantitatively, exhibiting reduced artifacts and significantly improved performance (up to 11 %) in classification tasks. Mahmood et al. [65] investigated the role of stain normalization in facilitating object detection tasks in digital pathology images. They evaluated a nuclei detection algorithm on normalized versus non-normalized images from multiple staining protocols and scanners. Their findings illustrated that stain normalization notably enhances the segmentation performance of computer-assisted analysis techniques, showcasing an increase in the Dice score by up to 3.5 %. Alsubaie et al. [66] proposed a novel stain deconvolution method using statistical analysis of multi-resolution stain color representation. It separates stain colors from histological images by applying independent component analysis in the wavelet domain, achieving good stain separation without artifacts in normalized images compared to other methods. Zheng et al. [67] introduced an innovative adaptive color deconvolution model tailored for stain separation and normalization. Their model achieved notably more color-consistent normalization outcomes by integrating prior knowledge of staining alongside intensity constraints. Notably, this approach led to an impressive 7.2 % increase in the AUC when tested on an external dataset. Martos et al. [68] proposed a fully automated pipeline for nuclei segmentation in gastric cancer images. It performs color normalization using an optical density colorspace conversion. Their method notably enhances the F1-measure by 7.1 % when evaluated on external test sets. Wang et al. [69] proposed a Fourier-based mitosis detection method that tackles domain shift by using fast Fourier transformation on MIDOG 2021 challenge data. It replaces the low-frequency spectrum of source domain images

with that of a reference domain, generating new images through inverse FFT. This style transfer, enhances domain generalization without altering image details or labels, increasing the F1 score by 0.4 %. Aalhassan et al. [70] proposed the FFT-based data augmentation to enhance model generalization across multicenter data. The end-to-end segmentation strategy outperforms the state-of-the-art methods by approximately +6.5 %. Bazargani et al. [71] proposed an innovative approach to enhance model robustness, departing from conventional methods like random augmentation. Their method aligns the H&E color space of the source dataset with both datasets, incorporating random color augmentation for a broader color distribution. This strategy significantly improves generalization, reflected in an increased AUC of 0.03 and 0.05 for internal and external datasets, respectively.

Subsequently, research moved to more sophisticated techniques based on machine learning. These techniques use classifiers to recognize stains within the image and consequently normalize them. One such approach employs sparse coding to separate blended stains into their components for normalization [72]. Perez-Bueno et al. [73] introduced a framework for blind color deconvolution, normalization, and classification of histological images. The method combines Bayesian modeling and inference with sparse priors to separate multi-stained images into single-stained components. It then normalizes the images, and experimental evaluations show that the proposed approach outperforms state-of-the-art methods in terms of preserving tissue structure and enhancing cancer classification accuracy by 1.6 % in terms of AUC. Bejnordi et al. [74] presented the Whole-slide Image Color Standardizer (WSICS), which classifies pixels into stains, transforms stain distributions to a template, and combines transformations with weights. WSICS performs superior compared to baseline methods in normalization tasks, showcasing a noTable 5.5 % enhancement in AUC performance specifically observed in rat liver images. Additionally, it displays improved color constancy in lymph node images, emphasizing its efficacy across varied tissue types. Khan et al. [75] presented a nonlinear mapping approach to stain normalization in digital images. The proposed method employs a spline-based nonlinear transformation of channel statistics to normalize color variations introduced during the staining process. The results demonstrate the effectiveness of the approach in preserving image structure and enhancing visual quality, making it suitable for computer-aided image analysis in histopathology. Shafei et al. [76] introduced a novel approach called "Class-Agnostic Weighted Normalization" (CLAW normalization) for stain normalization in histopathology images. The method utilizes a mixture of multivariate skew-normal distributions for stain clustering and parameter estimation, combined with a stain transformation technique. The results show that the proposed approach outperforms existing methods in terms of information preservation, enhancing visual quality, and improving classification accuracy by 7 % when compared to original images.

More recently, methods based on deep learning have become increasingly common. These techniques use deep neural networks or generative models to normalize images without necessarily decomposing the stains. Deep learning approaches have demonstrated an ability to learn complex patterns in histology images and perform stain normalization in an end-to-end manner [77]. Janowczyk et al. [78] presented StaNoSA, a deep learning-based approach for stain normalization. It uses sparse autoencoders to perform unsupervised tissue partitioning, allowing histogram matching to be done on a per-tissue basis. This achieves comparable or better normalization compared to other techniques, managing variability from staining, scanning equipment, and tissue class imbalance. Zaneta Swiderska-Chadaj et al. [79] discussed the impact of rescanning, stain normalization, and their combination on the performance of convolutional neural networks in the multi-centric, whole-slide classification of prostate cancer. Their evaluation found that combining rescanning and normalization techniques improves CNN accuracy by up to 10 % in classifying prostate cancer on whole-slide images. Perez et al. [73] presented a deep learning solution based on contrastive learning to transfer between different staining styles. They evaluated the model on two digitized datasets and achieved an improved classification accuracy of up to 13 % compared to unprocessed images. Kang et al. [54] presented StainNet, a stain normalization network that utilizes pixel-wise adjustment via a fully convolutional network. By employing distillation learning, StainNet achieves performance comparable to deep learning methods while better preserving source information and operating more than 40 times faster than previous methods. Marini et al. [80] introduced a CNN that learns stain-invariant features through regression. By learning from paired stained and unstained images, the CNN can focus on underlying tissue characteristics rather than on color patterns alone. Tellez et al. [81] proposed a stain normalization method using a U-Net architecture that maps augmented stain versions to a normalized representation. Compared to prior methods, it achieved significantly better generalization across organs and cancer types in computational pathology applications. Sun et al. [82] proposed Deep Attention Integrated Networks (DAINets) for nucleus segmentation. It designs an Individual Color Normalization strategy to address stain variation issues across multi-organ images. Evaluations demonstrated it achieves state-of-the-art performance on nucleus segmentation, showcasing an improvement of 1.7 % in the Dice score compared to the unprocessed image. Jeong et al. [83] developed a score-based diffusion model with stain separation and overlapping patches for stain normalization. Their approach involved decomposing and normalizing of individual stains using a diffusion model. The evaluation of colon biopsy images demonstrated that the method achieved high-performance stain normalization, with a Pearson correlation coefficient of 99 %. Additionally, the normalization prevented artifacts from being introduced to the images during the process.

One of the most common approaches for stain normalization of histopathology images involves adversarial training and generative adversarial networks (GANs). GANs can be used to learn the relationship between stained and unstained versions of images and translate between these two domains. Bentaieb et al. [84] proposed an adversarial deep learning approach using a GAN to learn the mapping between stained and unstained images, enabling stain transfer between images. Experimental results demonstrated that the proposed method improved the performance of image analysis tasks such as mitosis detection (+18 % accuracy), colon cancer classification (+5.1 % accuracy), and ovarian cancer classification (+16.9 % accuracy). Shrivastava et al. [85] introduced a self-attentive adversarial approach for stain normalization, utilizing self-attention and adversarial training to normalize multiple stain domains to a common domain while preserving cellular structures. Lafarge et al. [86] presented a domain-adversarial framework for learning domain-invariant representations, achieving improved generalization over conventional methods in mitosis detection and nuclei segmentation tasks. Salehi et al. [87] proposed a Pix2Pix-based stain translation method to address inconsistent staining by translating between staining protocols. Experimental results demonstrate that the proposed method effectively normalizes stains and improves the performance of downstream analysis tasks such as nuclei segmentation (up to 3.5 % F1-measure) and mitosis detection (+30 % accuracy). Shaban et al. [88] presented a CycleGAN-inspired solution that eliminates the need for expert template selection, testing on a clinical use case with a 12 % increase in AUC over unprocessed images. Cong et al. [89] proposed a texture-enhanced generative adversarial network (TESGAN) for stain normalization using higher-contrast hematoxylin components as input to generate normalized images without reference images. Cong et al. [90] introduced a semi-supervised CAGAN approach using color augmentation and a dual-decoder GAN with consistency regularization. Their method, by learning from unlabeled source domain images, achieved a notable improvement between 5 % and 10 % in classification performance compared to established baseline methods.

Some comparative studies assess the impact of color normalization approaches in the AI framework. Boschman et al. [91] systematically investigated eight color normalization algorithms for AI-based classification of H&E-stained histopathology slides in the context of using

images from both one center and from multiple centers. Their results show that color normalization does not consistently improve classification performance when both the training and testing data are from a single center. However, using four multi-centric datasets of two cancer types (ovarian and pleural) and objective functions, they demonstrate that color normalization can significantly improve the classification accuracy of images from external datasets (ovarian cancer: 0.25 AUC increase; pleural cancer: 0.21 AUC increase). Altini et al. [92] investigated using of Unpaired Image-to-Image Translation (UI2IT) models for stain color normalization in histology images of colon cancer. The authors compare five DL normalization models based on GANs and propose a meta-domain training approach to reduce training time. The results show that the UI2IT frameworks provide realistically colorized images, improving the accuracy of downstream classifiers by up to 3.6 % compared to traditional normalization methods.

Some works in digital pathology utilize color augmentation strategies, which aim to increase the variability of the data to cover all possible color nuances rather than normalizing color variations. By expanding the diversity of training examples in this way, models can be made more robust to variations inherently present across whole slide images from different systems and staining protocols. Faryna et al. [93] presented a color augmentation strategy for histopathology images stained with H&E. The authors propose adapting traditional data augmentation techniques to the specific characteristics of histopathology images. This approach improves the performance of deep learning models on tasks such as classification, showcasing a remarkable increase in the AUC of 50.8 %. In another study, Faryna et al. [94] investigated the potential of automated hyper-parameter search for augmentation, aiming to enhance generalization in histopathology. They assessed four advanced automatic augmentation methods across 25 centers for tumor metastasis detection and breast cancer tissue classification. Results reveal comparable performance in metastasis detection and a significant outperformance in breast cancer classification compared to manual augmentation. Marini et al. [95] introduced Data-Driven Color Augmentation (DDCA), a method to enhance color augmentation by comparing augmented stain matrices to a database of variations. DDCA was applied to color augmentation and adversarial training, outperforming baseline methods in classifying colon and prostate images. It showed better generalization to heterogeneous data, improving up to 26.7 % in classification performance. Demmaka et al. [96] explored the application of color augmentation to enhance the detection of tumor mutational burden in scans of H&E-stained multicenter slides of lung squamous cell carcinoma. By implementing augmentations like random brightness transforms and random whole-color-channel pixel intensity shifts, they observed a notable AUC increase from 0.70 to 0.90. Huang et al. [97] introduced the OmniCE augmentation method, and experimental findings demonstrate its superior performance compared to Augmix augmentation. The model trained on OmniCE-augmented datasets outperforms Augmix-augmented ones by 8.3 % and 15.3 % at two distinct centers, achieving state-of-the-art (SOTA) performance. Otalora et al. [98] compared stain normalization, color augmentation, and domain adversarial training approaches to improve the generalization of classification networks to external datasets. Results showed that color augmentation and stain normalization achieved the best generalization by learning stain-invariant representations of tissue images. Bouteldja et al. [99] investigated different approaches to improve the generalization of a pretrained kidney segmentation CNN to external cohorts with distinct stain variations. They proposed augmenting the training data with external stain variability using CycleGANs and comparing this with stain normalization approaches. Their proposed stain augmentation approach outperformed other methods at segmenting kidney structures in external cohorts, yielding a significantly improving in the Dice score from 1.2 % to 2.1 %. Tolkach et al. [100] introduced a deep learning model designed for detecting tumor tissues and grading histological regression in esophageal adenocarcinomas. Their preprocessing approach suggested augmentation and stain normalization, incorporating both Macenko stain transfer and CycleGAN-based transfer. Notably, using CycleGAN in tissue regression led to a remarkable increase of +0.12 in AUROC in external cohorts.

Despite the excellent performance achieved by color augmentation techniques, it is important to consider the limitations of data augmentation in providing all possible stain variations. Given the vast range of potential color variations that can occur in digital pathology images, it is practically impossible to include every single variation through data augmentation alone. This limitation highlights the need for image harmonization, also known as color normalization, as a more promising approach. Image harmonization ensures consistent color representation across different images and staining protocols, which can enhance model generalization. By harmonizing images, we can address the challenge of model generalization by reducing the influence of color variations and ensuring that the model focuses on relevant features and patterns in the pathology images.

Image harmonization techniques in digital pathology aim to standardize histopathology images and reduce variability caused by different staining protocols and imaging systems. Early approaches used color deconvolution and mathematical methods for normalization, such as histogram-based methods that increased classification accuracy by 13 % compared to unprocessed images [62] and real-time stain normalization systems that boosted classification accuracy by 5 % [63]. More sophisticated techniques based on machine learning, such as sparse coding and Bayesian modeling, further improved performance, exhibiting a 5.5 % enhancement in AUC performance [74] and improving classification accuracy by 7 % [76]. Recent works leverage deep learning, including GANs, to translate between stained and unstained images. These approaches have demonstrated significant improvements in various tasks, such as mitosis detection (+18 % accuracy), colon cancer classification (+5.1 % accuracy), and ovarian cancer classification (+16.9 % accuracy) [84]. Comparative studies have shown that color normalization can significantly improve the classification accuracy of images from external datasets, with increases in AUC ranging from 0.21 to 0.25 [91]. Data augmentation is also explored to enhance model robustness to staining variations, ensuring reliable results with different scanners. These techniques have led to remarkable improvements in classification performance, with increases in AUC ranging from 26.7 % [95] to 50.8 % [93]. While the creation of slides may retain a manual aspect despite automation efforts, digital normalization becomes pivotal. It ensures uniform coloration and slide quality, not just for AI but also for pathologists. This guarantees swift and reliable diagnoses, an essential factor for validation across varied pathology labs and imaging systems in multi-site studies.

3.3.2. *Fluorescence imaging*

Table A6 summarizes the works discussed in this section, along with the impact of the strategies employed. Among the techniques employed for image harmonization in fluorescence imaging, denoising and grayscale normalization emerge as the most commonly used methods (Fig. 9). Denoising aims to improve the signal-to-noise ratio, further refining the visibility of crucial structures and enhancing overall image quality. For denoising, machine learning methods are primarily used, with one study showcasing a deep learning approach. On the other hand, grayscale normalization employs only mathematical methods. Notably, five out of the six studies examined in this section focused on analyzing image quality metrics, to standardize and improve image quality obtained from different facilities using various acquisition devices across multiple centers. Only one study deals with classification tasks.

In terms of image denoising, approaches have been presented based both on machine learning and deep learning. Yang et al. [101] proposed a noise reduction algorithm employing machine learning that estimates noise parameters through contourlet transform coefficients. Their experiments on fluorescence microscopy images demonstrated enhanced denoising performance compared. Specifically, their method achieved a notable improvement with a 2 dB margin in PSNR. Mannam et al. [102]

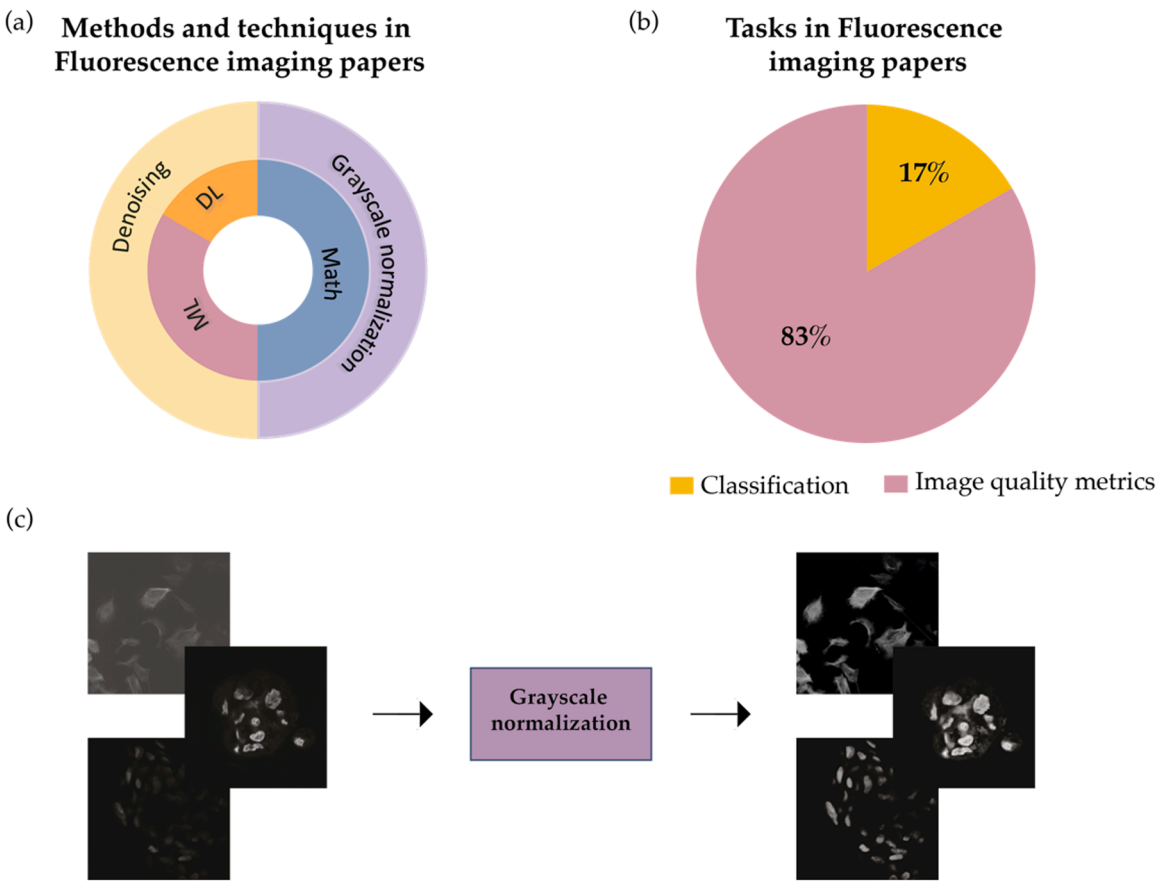

Fig. 9. Summary of the studies ($n = 6$) on image harmonization in fluorescence. (a) Techniques employed in fluorescence imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric fluorescence imaging studies. (c) Example of image harmonization in fluorescence imaging. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

introduced a CNN trained specifically to address denoising in fluorescence microscopy, particularly in mixed Poisson-Gaussian noise scenarios. External dataset evaluations demonstrated remarkable improvements, showing up to an 8 dB enhancement in PSNR when compared to analytical methods and machine learning approaches. The method was validated on various fluorescence samples with different noise types and contrast ratios. Yang et al. [103] introduced DeepNoise, a deep learning model tailored to disentangle biological signals from technical noises within high-content image-based assays. This model exhibited impressive accuracy, reaching 99.5 % in classifying 1108 distinct genetic perturbations screened across 125,510 fluorescent microscopy images. The successful isolation of biological and technical factors holds promise for reducing treatment development costs and expediting drug discovery processes. Broaddus et al. [104] presented STRUCTN2V, a method specifically designed for self-supervised denoising fluorescence microscopy images affected by structured noise. This technique employed blind spot networks, utilizing extended blind masks to conceal pixels and effectively eliminate spatially correlated noise. Across two real microscopy datasets, STRUCTN2V showcased superior performance compared to standard and blind spot techniques, showcasing enhancements in PSNR (+1.6 dB) and SSIM (up to 2.97 %) metrics. Zhang et al. [105] introduced the Fluorescence Microscopy Denoising (FMD) dataset, comprising 12,000 authentic fluorescence microscopy images afflicted by Poisson-Gaussian noise. This diverse dataset encompasses various microscopy modalities and encompasses representative biological samples. Through benchmarking 10 denoising algorithms, the authors discovered that deep learning methods exhibited superior performance. These methods showcased an impressive enhancement in PSNR by 10.97 dB and SSIM by 5.67 % when compared to the original, untreated images. Demircan-Tureyen et al. [106] proposed to tailor a dataset for training a denoising CNN for fluorescence microscopy images where ground truths are limited. Their approach involved leveraging low-level image features to curate visually similar images. They fine-tuned a pretrained CNN using a limited amount of target data. Remarkably, this approach showcased superior outcomes across two datasets compared to the original images, marking a substantial increase in PSNR from 4 dB to 9.6 dB and enhancing SSIM by up to 37.3 %.

In summary, image harmonization—primarily through denoising techniques—is a crucial role in fluorescence imaging studies. It effectively addresses challenges arising from different acquisition devices, variations in protocols, and the absence of standardized analysis methods, ensuring reliable and valid research findings. The studies discussed here underscore the significant impact of denoising methods, utilizing both machine learning and deep learning approaches, in improving image quality and ensuring consistent sample representation across various facilities and acquisition devices. Machine learning-based denoising approaches, such as the one proposed by Yang et al. [101], have demonstrated enhanced performance compared to conventional techniques, achieving a notable improvement of 2 dB in PSNR. Deep learning-based methods have shown even more impressive results, with CNNs showcasing up to an 8 dB enhancement in PSNR when compared to analytical methods [102]. The STRUCTN2V method [104], designed for self-supervised denoising of fluorescence microscopy images affected by structured noise, outperformed standard and blind spot techniques, with enhancements in PSNR (+1.6 dB) and SSIM (up to 2.97 %) metrics. These advancements in denoising techniques have notably improved metrics like PSNR and SSIM, promising a more reliable and standardized

approach to fluorescence imaging in multi-centric studies.

3.3.3. OCT/OCTA imaging

Table A7 summarizes the works reported in this section along with the impact of the strategy employed while Fig. 10 shows the distribution of the methods, techniques, and tasks performed in the OCT/OCTA papers analyzed. Various techniques have been employed for image harmonization in OCT/OCTA studies. In particular, both normalization and denoising approaches are applied as well as contrast enhancement and image resampling methods. Normalization aims to account for differences in brightness/contrast levels across datasets acquired using varying OCT systems and protocols. Denoising helps reduce background speckle noise that can obscure anatomical features. Contrast enhancement further improves the visibility of structures by accentuating intensity variations. Resampling reformats raw image dimensions to a consistent resolution, rectifying discrepancies in pixel sizes and scan areas. The most commonly used methods are deep learning (for denoising and grayscale normalization) and mathematical methods (for grayscale normalization, resampling, and contrast enhancement). Additionally, segmentation was the predominant task examined (67 %), followed only by quality metric studies.

One of the main challenges when analyzing an OCT image is speckle noise, making a correct interpretation of the image difficult for experts and CAD systems. Shi et al. [107] proposed DeSpecNet, a CNN for OCT retinal image despeckling. The method exhibited strong generalization capabilities across four scanners with different wavelengths. Specifically, despeckled images saw an improvement of approximately 14 % PSNR and approximately 5.3 % on CNR compared to the original speckled images. For speckle denoising, Gour et al. [108] presented a residual CNN that adapted the pre-trained VGGNet architecture. The proposed method's strength was its ability to generalize a speckle denoising model across different images. It demonstrated an improvement of approximately 10 dB in PSNR and approximately 60 % in structural similarity index measure (SSIM) compared to the baseline speckled images.

Image harmonization can also address the limitations of existing segmentation methods, which often have restricted applicability to samples closely resembling the distribution of the training data. Romo-Bucheli et al. [109] proposed an unsupervised unpaired image translation approach using CycleGANs to address variability between images from distinct OCT devices. CycleGANs were employed due to their effectiveness in handling variability observed across different OCT scanners. The results demonstrated that applying the translation algorithm significantly improved segmentation model performance when classifying images from a different vendor. Specifically, the Dice score improved by 47 % for intra-retinal cyst segmentation, 54 % for sub-retinal fluid segmentation, and 29 % for photoreceptor layer segmentation in the test set. Venhuizen et al. [110] aimed to automatically segment the inner retinal complex (IRC) in OCT images acquired from different devices. As a pre-processing step, they applied resampling to standardize the spatial resolution of the OCT scans. Then, they implemented a cascade of two fully convolutional neural networks (FCNNs) to segment the IRC. Bogunovic et al. [111] described eight ways to detect and segment the three retinal fluids in a multi-centric OCT study using various segmentation algorithms and pre-processing methods. The works described in the paper used image harmonization techniques like median filtering, morphological operators and 3D smoothing. The top-performing method applied 3D bounded variation denoising to

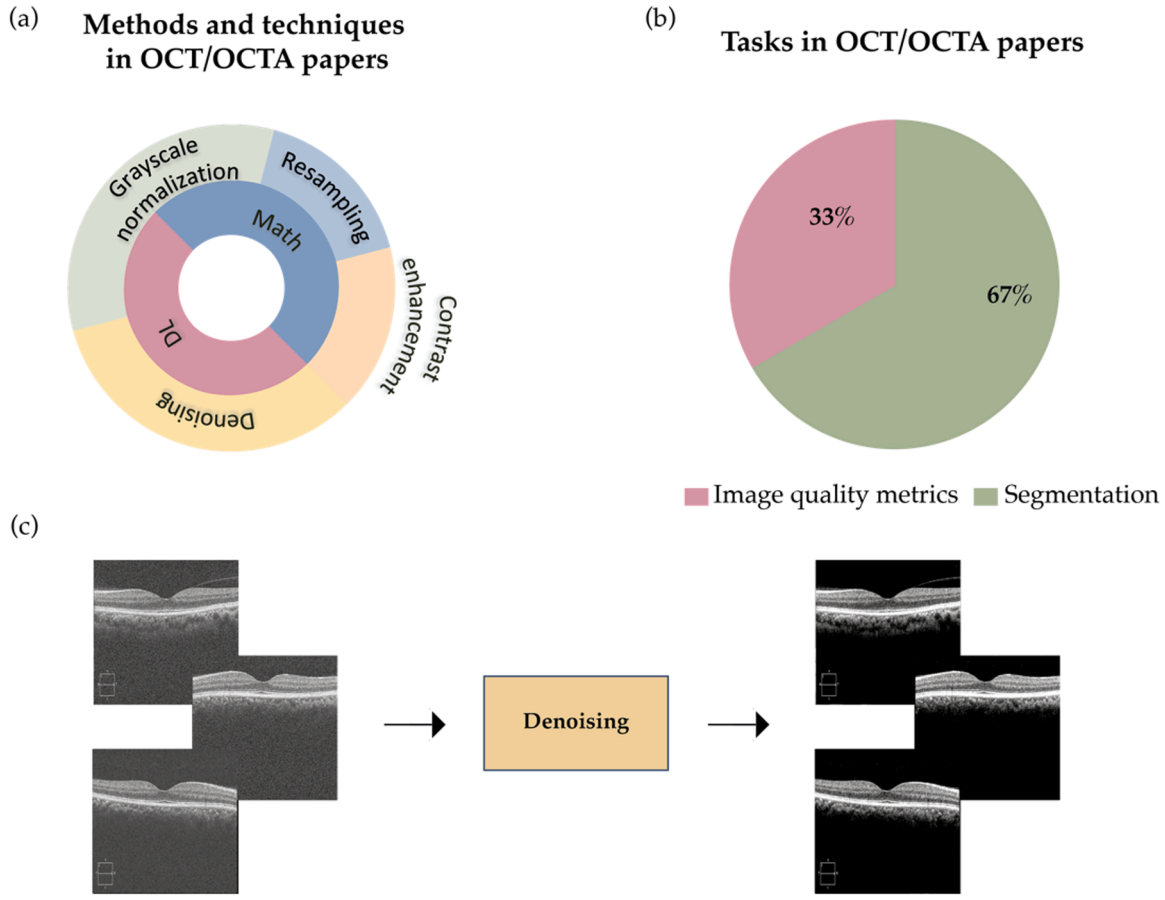

Fig. 10. Summary of the studies ($n = 6$) on image harmonization in OCT/OCTA. (a) Techniques employed in OCT/OCTA imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric OCT/OCTA imaging studies. (c) Example of image harmonization in OCT/OCTA. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

motion-corrected B-scans, achieving a mean Dice score for intra-retinal fluid, sub-retinal fluid, and pigment epithelial detachment of 82 %, 75 % and 74 % respectively.

Regarding OCTA, the only multi-device study that adopted image harmonization was conducted by Ma et al. [112]. They utilized contrast-constrained adaptive histogram equalization (CLANE) prior to segmenting the complex vascular network in the images. This multi-device study achieved a 0.4 % improvement in the Dice score after applying CLANE.

Image harmonization techniques show great promise for advancing clinical research using OCT/OCTA modalities. A key challenge is variability between imaging systems and acquisition protocols across research sites, which can hamper efforts to pool and jointly analyze large multi-centric datasets. Normalization, denoising, contrast enhancement and resampling approaches have been applied to standardize brightness, reduce speckle noise, improve structure visibility, and rectify pixel mismatches between OCT datasets, allowing for more robust cross-site comparability. In terms of denoising, the despeckle of OCT retinal image exhibited strong generalization capabilities across four scanners with different wavelengths, with an improvement of 14 % in PSNR and 5.3 % in CNR compared to the original images [107]. Similarly, a residual CNN demonstrated an improvement of approximately 10 dB in PSNR and 60 % in SSIM compared to the baseline speckled images [108]. Generative models have also been employed to address the limitations of existing segmentation methods. A CycleGAN for unsupervised unpaired image translation was developed to handle variability between images from distinct OCT devices. The results showed significant improvements up to 54 % in Dice score [109]. Overall, image harmonization addresses critical heterogeneity issues, enabling larger and more diverse OCT/OCTA cohorts that can help segmentation and diagnostic deep learning algorithms achieve their full potential for advancing clinical care.

3.4. Ultrasound imaging

Table A8 summarizes the works reported in this section and the impact of the strategy employed. As shown in Fig. 11, most studies analyzed focused on grayscale normalization of US images. The primary objective of these harmonization techniques was to enable image classification using data pooled from different scanners and operators. Only one study utilized contrast enhancement for harmonization, based on mathematical methods. The analyzed studies in literature all employed mathematical algorithms to standardize ultrasound image appearance before multi-center analysis. Classification was the most studied downstream task (83 %), followed by segmentation. Fig. 11 also provides an example demonstrating the effects of grayscale normalization on ultrasound images from different centers, highlighting its ability to improve harmonization for multi-center US data.

Liu et al. [113] aimed to quantify the risk level of gastrointestinal stromal tumors in multi-centric endoscopic ultrasound images. They introduced a triple normalization approach to address issues from multi-centric data bias. Their harmonization approach included applying a CLAHE algorithm and resizing the images. An ablation experiment was conducted to further investigate the impact of this intensity, size, and spatial resolution normalization, demonstrating a 3 % increase in classification accuracy. In another work, Ren et al. [114]

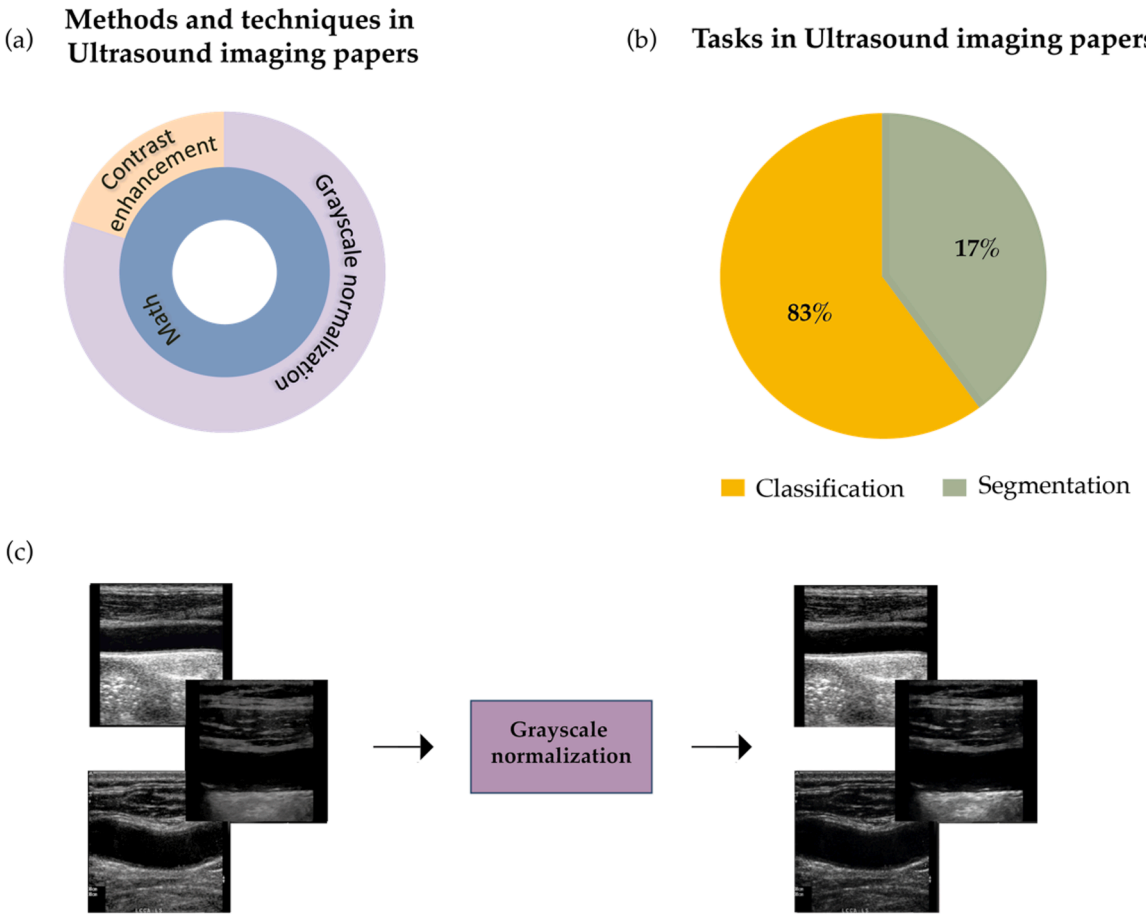

Fig. 11. Summary of the studies ($n = 5$) on image harmonization in ultrasound. (a) Techniques employed in ultrasound imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric ultrasound imaging studies. (c) Example of image harmonization in ultrasound imaging. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

aimed to grade hepatocellular carcinoma using machine learning-based techniques on data from three hospitals. The authors employed z-score normalization, resampling through B-spline interpolation and gray-level discretization to standardize the appearance of the ultrasound images, obtaining an AUC of 0.874. Homayoun et al. [115] proposed a 3σ normalization approach to mitigate undesired consequences arising from the utilization of diverse ultrasound devices across the three centers to classify breast lesions using ultrasound radiomics signatures. Sirjani et al. [116] implemented min-max scaling to normalize the grayscale of ultrasound images from 5 different datasets. The final goal of classifying breast lesions was achieved with an accuracy of 81 % on an external test set, demonstrating higher performance compared to than other popular CNN architectures. Specifically, VGG attained an accuracy of 75 %, DenseNet achieved 73 %, and ResNet achieved 67 % for the task of classifying breast lesions. The only study that employed image harmonization for segmentation was performed by Du et al. [117]. They proposed a multi-centric study for segmenting intravascular ultrasound images using convolutional neural networks. The study included 175 intravascular ultrasound pullbacks obtained from two medical centers. To avoid biases, z-score normalization was performed on all the images. The proposed method achieved a Dice score of 92.7 %.

The studies reported in this section demonstrated a variety of techniques can help standardize the appearance and enable fair comparisons across ultrasound data. Key approaches included intensity normalization methods like histogram equalization, z-score normalization, and min-max scaling to mitigate differences in grayscale values and contrast. Spatial normalization through resizing and interpolation facilitated input uniformity for classification and segmentation models. Intensity, size, and spatial resolution normalization led to a 3 % increase in classification accuracy for gastrointestinal stromal tumor risk assessment [113] and improved the AUC up to 0.874 for grading hepatocellular carcinoma [114]. In a study applying z-score normalization to images from two centers, a Dice score of 92.7 % was achieved for the segmentation of intravascular ultrasound images [117]. These quantitative results highlight the significant impact of harmonization techniques on downstream tasks in multi-centric ultrasound studies. However, the limited number of studies and the focus on a narrow range of tasks (primarily classification) suggest that further research is needed to understand the potential and limitations of these methods fully. Continued development of such harmonization methods is crucial to fully leverage ultrasound's significant potential for collaborative, multi-centric research, and clinical applications.

3.5. Dermatology imaging

Table A9 summarizes the works reported in this section along with the impact of the strategy employed. Harmonization techniques in dermatology encompass various aspects, such as standardizing imaging parameters, calibrating devices, and optimizing lighting conditions. These approaches aim to reduce variations caused by different equipment and settings, thereby enabling more accurate and comparable analysis of dermatological data. As shown in Fig. 12, most works use color normalization to standardize the appearance of dermatological images. Color normalization techniques have been applied to both classification and segmentation tasks. Mathematical methods are the most commonly used for color normalization and resampling, followed by deep learning approaches. In summary, harmonization research in dermatology focuses on reducing equipment and environmental

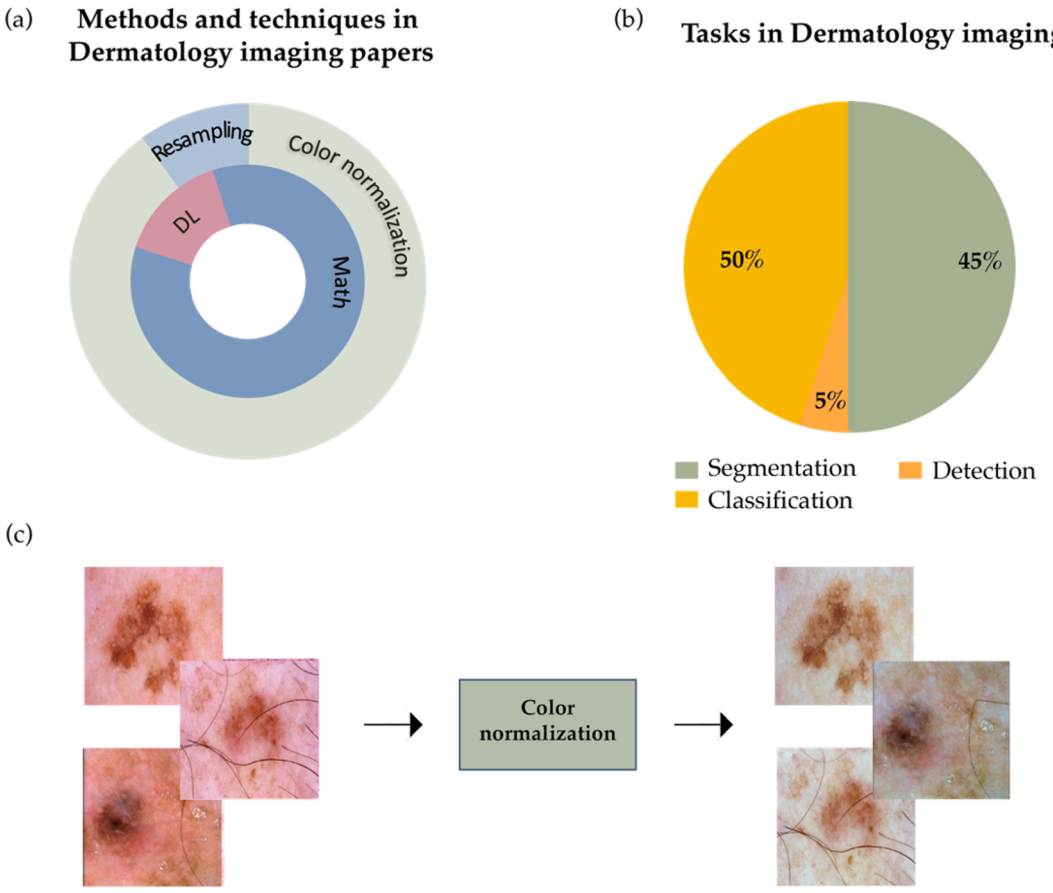

Fig. 12. Summary of the studies ($n = 19$) on image harmonization in dermatology. (a) Techniques employed in dermatology imaging categorized as either machine learning (ML), deep learning (DL), or mathematical methods. (b) Distribution of tasks in multi-centric dermatology imaging studies. (c) Example of image harmonization in dermatology. The inner circle of (a) shows the type of technique used in the approach described by the outer circle of the same graph.

differences through techniques like color normalization, facilitating multi-site analysis.

The early approaches to image harmonization in dermatology were simple techniques such as scaling pixel values using min-max scaling and z-score normalization strategies. Codella et al. [118] introduced a system integrating deep learning, sparse coding, and support vector machines (SVMs) with multi-contextual analysis for melanoma segmentation and classification. They cropped and resized dermatological images before feeding them into the DL network. Their results showed an improvement of 7.5 %, achieving an AUC of 0.843 for classification and a Jaccard index of 84 % for segmentation. Azad et al. [119] introduced TransNorm, a transformer-based model designed for medical image segmentation. Their approach involved resizing the images to a fixed size and integrating a spatial normalization mechanism within the transformer module to adaptively recalibrate skip connections. Through evaluation across multiple datasets, their method showcased considerable effectiveness, enhancing segmentation performance by up to 8 % in the Dice score when compared to other network architectures. Yu et al. [120] introduced a framework for classifying dermoscopy images that leverages deep convolutional features and Fisher vector encoding. To normalize images, they applied a process of subtracting the per-image-mean from each channel. Evaluating the ISIC 2016 dataset, their method demonstrated superior performance compared to state-of-the-art methods, enhancing the mean average precision (mAP) by approximately 1.3–6 %. Gong et al. [121] introduced a decision fusion technique for classifying dermoscopy images employing multiple pre-trained CNNs. Their image normalization process involved subtracting the channel-wise average intensity from each image. Evaluating the ISIC 2019 dataset, the decision fusion approach showcased enhanced performance compared to individual CNNs and traditional fusion methods. It achieved an accuracy of over 99.5 % and a specificity of 99.6 %. Shahin et al. [122] introduced a DCNN model to classify skin lesions as benign or malignant. Their approach involves data normalization through z-score normalization and employs data augmentation to expand the training dataset. Their model achieved an impressive test accuracy of 91.93 % on HAM10000, surpassing the performance of transfer learning models such as ResNet, VGG-16, and MobileNet. Xin et al. [123] introduced the SkinTrans model, leveraging a vision transformer for skin cancer classification. Their approach involves multi-scale patch extraction from images and utilizes contrastive learning to encode similar data. They employed z-score normalization for the input images. Their model demonstrated a noteworthy increase of 1 % in accuracy compared to the baseline. Gajera [124] analyzed dermoscopy images for melanoma detection utilizing deep CNN features. Following image resizing and cropping, they applied per-channel normalization using the z-score method. Employing an MLP classifier resulted in an impressive accuracy of 98.33 % for melanoma detection, showcasing up to a 5 % enhancement compared to other networks that did not employ any image harmonization step. Zafar et al. [125] introduced a skin lesion segmentation technique utilizing a Res-Unet convolutional neural network model. They conducted normalization to scale pixel values within the range of 0 to 1. Additionally, they implemented hair removal using morphological operations. Their method exhibited an enhancement in the Jaccard index of up to 2 % when compared to other existing approaches. Behara et al. [126] introduced a skin lesion classification framework with a focus on the qualitative assessment of the preprocessing phase, including techniques such as bicubic interpolation for scaling, normalization, sharpening, color transformation, and median filters. Their model demonstrated superior performance with a notable accuracy of 99.38 %, outperforming all the compared methods.

These initial methods focused on the basic standardization of input images to deep learning models by rescaling pixel intensity ranges linearly. While helping to some degree with training, they did not address more complex issues like variations in lighting, skin tone, image distortions, and background noise between images. More sophisticated preprocessing involving color constancy algorithms would later be developed to better normalize dermoscopic images prior to analysis. Barata et al. [127] aimed to address issues arising from color variations in dermoscopy images acquired from different sources. They investigated employing color constancy techniques based on shades of gray to execute color normalization. Their study showcased enhanced performance, exhibiting an increase of up to 14 % in accuracy for two classification systems when color normalization was applied. This highlighted the technique's efficacy in mitigating the impact of variations in acquisition setups. Barata et al. [128] explored color constancy algorithms aimed at standardizing the colors of dermoscopy images collected from various sources. They implemented four distinct color constancy methods and evaluated their impact on a bag-of-features classification system. Their findings revealed notable improvements: when leveraging the Shades of Gray color constancy method, sensitivity increased from 71 % to 79.7 % and specificity rose from 55.2 % to 76.8 %. Abbas et al. [129] introduced a melanoma border detection technique employing color normalization and region segmentation. Their method involved normalizing dermoscopy images to the CIE Lab* color space, enhancing contrast, and subsequently identifying regions of interest using a hill-climbing approach. Through experimental evaluations conducted on 100 images, their method achieved promising results with a true detection rate of 94.25 % and a false positive rate of 3.56 %, surpassing the performance of other existing methods in this domain. Ng et al. [130] delved into the impact of color constancy algorithms on the semantic segmentation of skin lesions. They utilized four distinct color constancy algorithms to preprocess images sourced from the ISIC Challenge 2017 dataset before training a Fully Convolutional Network (FCN) for segmentation purposes. Their findings indicated that preprocessing images with color constancy algorithms led to improved segmentation outcomes, particularly for seborrheic keratosis lesions, showcasing an enhancement of up to a 4 % increase in Jaccard similarity. Olga et al. [131] investigated the performance of an automatic lesion classification algorithm on skin cancer detection with and without image enhancement. They applied various enhancement methods, including the Retinex method [108], and conducted classification using CNNs and SVMs. Their findings revealed that the Retinex method yielded the most impressive performance, enhancing the F1 score by up to 5 % compared to scenarios with no preprocessing. Zhang et al. [132] introduced an attention residual learning convolutional neural network (ARL-CNN) designed for skin lesion classification. They incorporated the gray-world color constancy algorithm [133] as a preprocessing step before feeding the images into their network. Specifically for melanoma classification, the integration of color constancy with their novel network architecture increased the AUC by up to 14 %. Yuan et al. [134] proposed an improved Convolutional-Deconvolutional Network (CDNN) specifically for skin lesion segmentation. All images were resized to keep a balanced aspect ratio while reducing computational cost. In addition to the RGB color channels, three channels from the HSV color space and one channel (L) from the CIELAB color space were included as input to the network, resulting in a total of 7 channels. Remarkably, their proposed CDNN method secured the top position on the ISBI 2017 skin lesion segmentation challenge dataset, achieving an average Jaccard Index of 76 % on the testing set. Goyal et al. [135] introduced ensemble deep learning techniques for skin lesion segmentation using Mask R-CNN and DeeplabV3+. Their approach incorporated a color constancy step that followed the shades of gray algorithm [136]. The ensemble methods achieved higher sensitivity and specificity than other algorithms, showcasing an improvement in sensitivity from 4.4 % to 22.7 %.

More recently, advanced approaches leveraging deep learning have been presented for performing color normalization. Where earlier methods relied on traditional computer vision techniques, newer techniques utilize neural networks to learn complex mappings between image appearances under different lighting conditions in an end-to-end fashion. This allows for non-linear color normalization without requiring explicit modeling of the imaging pipeline. Salvi et al. [137]

introduced the DermoCC-GAN method for standardizing dermatological images via generative adversarial networks (GANs). They employed a custom heuristic algorithm for color constancy, mitigating illumination variability during GAN training. DermoCC-GAN exhibited superior performance in both classification and segmentation tasks compared to alternative color constancy methods. In classification, it showcased a 3 % accuracy improvement, while in segmentation tasks, it demonstrated enhancements of up to 19 % in Dice score compared to the original images.

As seen in digital pathology, algorithms for color augmentation have also been proposed in dermatology to make networks less sensitive to variations in the illumination conditions of images. These techniques aimed to expand dermatology datasets by artificially modifying aspects like brightness, contrast, and color tone through data augmentation. Galdran et al. [138] proposed a data augmentation technique for skin lesion analysis using color constancy. It estimates illuminants from training images and then applies random illuminants for augmentation. Networks trained with this method achieve promising segmentation and classification results on a validation set for skin lesion analysis tasks. Specifically, a Dice score of 84.6 % was achieved for segmentation and an AUC of 0.873 was obtained for classification, indicating excellent discriminability of lesion types.

Image harmonization in dermatology ensures for ensuring reliable and comparable research outcomes and clinical applications. By standardizing the acquisition and preprocessing of dermatoscopic images across multiple centers, image harmonization allows for effective comparison and combination of datasets. Harmonizing data improves the accuracy and reliability of image analysis algorithms by mitigating variations caused by lighting conditions and imaging devices [139]. Color normalization techniques, such as min-max scaling and z-score normalization, improved classification accuracy by up to 7.5 % and increased the Jaccard index to 84 % for segmentation [118]. Spatial normalization through resizing and interpolation enhanced segmentation performance by up to 8 % in the Dice score compared to other network architectures [119]. Z-score normalization contributed to an increase of 1 % in accuracy compared to the baseline [123]. It showcased up to a 5 % enhancement in accuracy compared to other networks that did not employ any image harmonization step [124]. Color constancy techniques based on shades of gray exhibited an increase of up to 14 % in accuracy for two classification systems [127]. The Shades of Gray color constancy method increased sensitivity from 71 % to 79.7 % and specificity from 55.2 % to 76.8 % [128]. Preprocessing images with color constancy algorithms led to improved segmentation outcomes, particularly for seborrheic keratosis lesions, showcasing an enhancement of up to a 4 % increase in Jaccard similarity [130]. The integration of color constancy with a novel classification architecture resulted in an increase in the AUC by up to 14 % [132]. An ensemble method incorporating a color constancy step following the shades of gray algorithm improved sensitivity from 4.4 % to 22.7 % [135]. Overall, image harmonization techniques play a crucial role in advancing dermatology by improving research outcomes, diagnostic accuracy, and the quality of patient care [140,141].

4. Discussion

4.1. Summary of main findings

Initially, AI tools were designed to work with data from a single center or a single acquisition. However, there has been a shift towards multi-centric and multi-device approaches over time. This shift is due to the limitations of working solely with single-center data, which can result in limited generalization, bias, and a drop in performance on the test set. Our review provides a comprehensive overview of image harmonization approaches in multicenter studies, analyzing the papers published between 2013 and 2023. Unlike previous reviews, which focused on normalization in one or a few imaging modalities, such as digital pathology [9,14] or radiology [10–13], our review covers a vast majority of the imaging modalities used in healthcare. Finally, our review emphasizes the impact of image harmonization on the performance of AI models in multicenter studies.

Fig. 13 analyzes an analysis of the number of papers that use image harmonization categorized by statistical/mathematical methods (Math), ML, and DL over the publication years. The figure reveals that mathematical models are the predominant approach consistently employed throughout the entire time frame. Statistical/mathematical techniques have been more widely utilized than DL or ML techniques in most years from 2013 to 2023. This preference for mathematical and statistical methods can be attributed to their inherent simplicity and interpretability. On the other hand, there has been a noticeable growth in DL-based methods since 2017. This surge in the use of DL can be attributed to the limitations of purely mathematical or ML-based approaches in managing image complexities, especially in color images, and handling variability introduced by changes in the acquisition system.

Recently, some studies have adopted standardization methods that combine mathematical models with DL or ML techniques, or even a combination of all three. For example, Altini et al. 2023 [92] employed a comprehensive standardization approach integrating mathematical, ML, and DL techniques in their color normalization strategy for digital pathology, utilizing methods such as color deconvolution, sparse non-negative factorization (SNMF), and GANs.

Fig. 14 provides the annual distribution of research papers across various application fields including digital pathology, dermatology imaging, fluorescence, mammography, PET/SPECT, MRI, ultrasound, CT, and OCT/OCTA imaging. This visualization offers valuable insights into the evolving trends and research interests within these application domains over time.

One notable finding from Fig. 14 is the increasing adoption of normalization techniques in the field of digital pathology, which has consistently remained a focal point of research, particularly in recent years (2019–2023). The chart also highlights the sustained application of normalization techniques in MRI and dermatology imaging throughout the analyzed period (2013–2023), with a noticeable increase in recent years. Additionally, novel normalization techniques have emerged for CT and OCT/OCTA imaging, indicating a shift in research focus that had not been explored extensively before 2018.

Fig. 15 provides insights into the utilization of image harmonization techniques across different application fields. It is noteworthy that mathematical and statistical models are the most employed techniques across almost all image modalities. The majority employed mathematical techniques for grayscale modalities, such as ultrasound and radiology imaging (CT, MR, PET/SPECT). The data further reveals that DL methods are more prominently used in imaging techniques characterized by greater image complexity. This includes fields like digital pathology, which deals with color images and significant variability in the

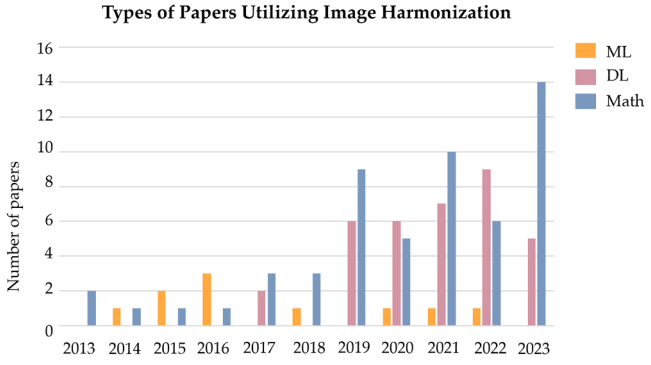

Fig. 13. Reviewed papers categorized by use of statistical/mathematical methods (Math), machine learning (ML), or deep learning (DL) approaches, divided by publication year.

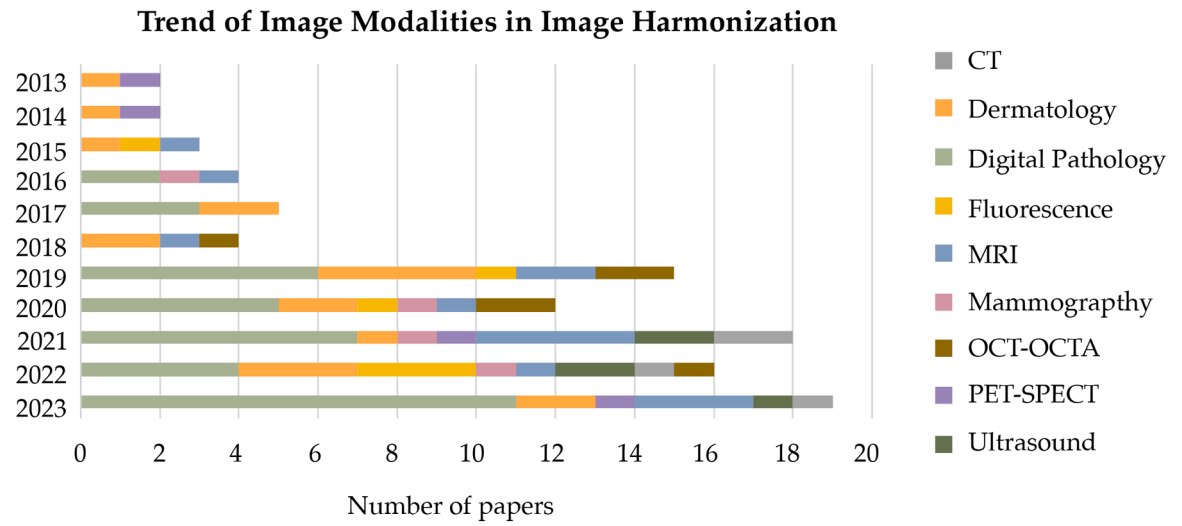

Fig. 14. The trends in the number of research papers categorized by major macroscopic application fields (radiology, nuclear medicine, optical imaging, ultrasound, dermatology) over the years.

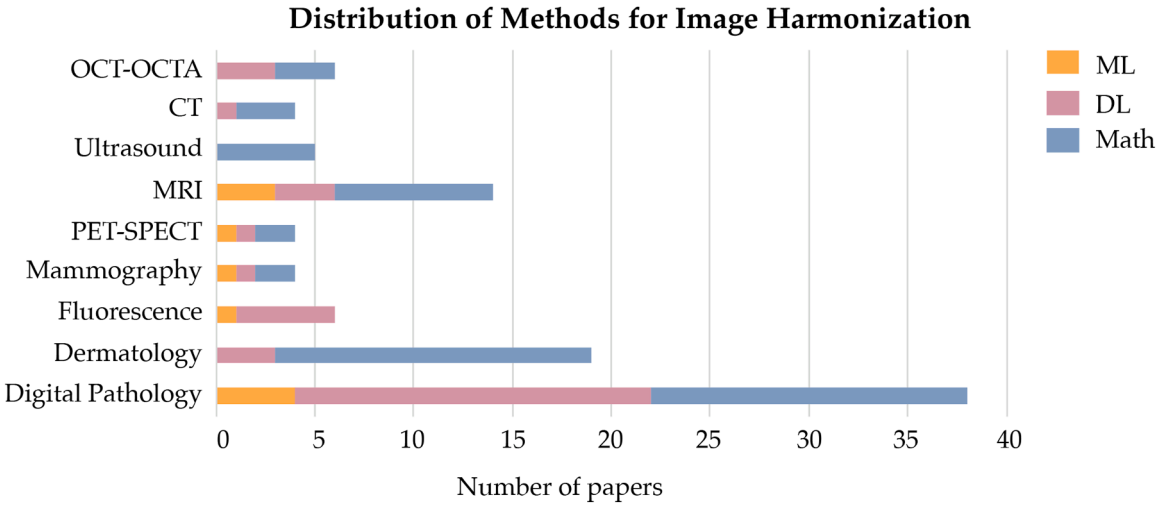

Fig. 15. Techniques of normalization (Math, ML, and DL) studies categorized by application area. Math represents the statistical and mathematics techniques.

sample preparation and acquisition system. DL methods are also increasingly utilized in emerging modalities like OCT/OCTA.

Fig. 16 clearly shows picture of the percentage of studies dedicated to image harmonization across different imaging modalities. Color image modalities like digital pathology (38 %, $n = 38$) and dermatology (19 %, $n = 19$) stand out as the most widely studied applications for multi-centric studies. This indicates the significance of harmonizing data collected from multiple centers in these fields. Furthermore, image harmonization is also a notable aspect in MR studies (14 %, $n = 14$), highlighting the importance of standardizing data collected from diverse sources in this particular application. Around 4 % of studies focus on image harmonization techniques in mammography and PET-SPECT. However, when it comes to ultrasound, CT, and OCT/OCTA, image normalization techniques have yet to become a predominant focus of research.

For each imaging modality, Table 1 presents a summary of the studies that demonstrate the most significant performance improvements using image harmonization techniques. The table highlights the authors, specific techniques employed, tasks addressed, and the quantitative impact of normalization on the results. The performance improvements are quantified using various metrics such as AUC, accuracy, PSNR, and SSIM, depending on the nature of the study and the task performed.

Notably, some recent studies introduce normalization techniques that apply to various fields, such as dermatology and digital pathology combined. Salvi et al. [141] introduced an application of GANs to tackle color variability in medical images. With a focus on digital pathology and dermatology, the method frames color normalization as an image-to-image translation problem, showcasing superior performance compared to existing methods in both domains. Table 2 shows the Open-access datasets used in multi-centric studies in healthcare.

4.2. Benefits and challenges of image harmonization for multi-centric studies

The integration of image harmonization has enormous potential to transform many areas of healthcare by providing a standardized and unified approach to data analysis in multi-centric studies. Adopting an image harmonization approach offers several key benefits:

- Improved robustness of the AI model towards out-of-distribution samples: image harmonization ensures that the AI model is robust on a diverse data, including samples from different centers. This improves the model's ability to generalize and perform well on

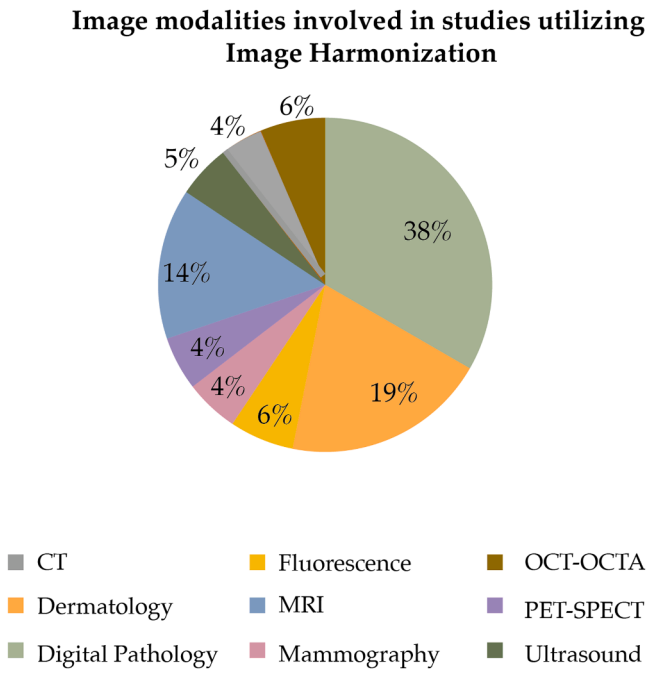

Fig. 16. Image modalities involved in the studies reviewed.

unseen data, enhancing its robustness and reliability in real-world applications [9].

- Bias reduction towards a specific medical center: by harmonizing data from multiple centers, bias towards any specific center or population can be minimized [158]. This helps to ensure that the AI model provides accurate and unbiased predictions across different medical settings, leading to fair and equitable healthcare outcomes.
- Multi-modal pipelines: normalization and harmonization of data across sources are essential steps to ensure the accuracy and reliability of a multi-modal system. Variability in how data is collected, including variations in image protocols, resolution, and image quality, can introduce systematic biases that compromise the effectiveness of multi-modal systems [159]. Data standardization pipelines should account for the inherent differences in data acquisition while preserving the biological variance present in the dataset.
- Facilitates collaborative research: image harmonization enables researchers from different institutions to collaborate more effectively. Establishing standardized data collection and harmonization protocols makes it easier to pool resources, share data, and conduct large-scale, multi-centric studies.
- Enables comparative analysis: harmonized data allows for direct comparisons between different centers, regions, or populations. This enables researchers to identify variations in disease prevalence, treatment outcomes, or other important factors, leading to improved understanding and more targeted interventions.

While integrating image harmonization has immense potential, several key challenges must be addressed to ensure its successful implementation:

- Reliable image harmonization: it is crucial to ensure that the data pre-processing and harmonization process does not lead to a loss of information or a decrease in the informative value of the processed image. The data preparation steps should be carefully designed and implemented to minimize any potential loss or distortion of important features in the original data.
- Variability in data formats and standards: Data generated by different imaging modalities or centers may have different formats, metadata structures, or standards. Addressing these variability issues is necessary to ensure interoperability and seamless integration of data across different sources. Standardization efforts are needed to establish common data formats and metadata standards to facilitate image harmonization and interoperability.
- Privacy and data protection: multi-centric studies involve sharing and integrating data from multiple institutions, raising privacy and data protection concerns. Safeguarding patient privacy and complying with ethical and legal requirements while sharing and harmonizing data is crucial. Data anonymization and secure data-sharing mechanisms need to be implemented to protect patient confidentiality.
- Data quality control and validation: Ensuring the quality and reliability of harmonized data is essential for accurate and meaningful analysis. Implementing rigorous data quality control measures and validation protocols is necessary to identify and address any data inconsistencies, errors, or artifacts that may arise during the harmonization process.

4.3. *Future research directions*

As the field of medical imaging continues to evolve, it is essential to highlight emerging trends and potential future directions in multi-centric approaches for healthcare. Several areas of improvement and future research opportunities exist to advance the field:

Table 1
Summary of studies highest performance improvements obtained for each imaging modality.

Author, year	Image modalities	Technique	Strategy	Task	Impact of normalization with respect to the baseline
Foltyn-Dumitru et al. [33]	MRI	Math	Grayscale normalization: z-score	Radiomic-based predictions for molecular glioma subtypes	AUC: 87 % (+42 %)
Tonneau et al. [48]	CT	Math	Resampling: voxel resampling, intensity clipping (HU), Denoising	Prediction of non-small cell lung cancer	AUC: 63 % (+11 %)
Thiele et al. [56]	PET/SPECT	Math	Grayscale normalization: intensity ratio	Classification of neurodegenerative dementias	Accuracy volumes: 88 % (+8 %). Accuracy scans: 86 % (+8 %)
Bejnordi et al. [74]	Digital pathology	ML	Color normalization: color deconvolution	Stain Specific Standardization of Whole-Slide Histopathological Images	AUC: 94.4 % (+63.4 %)
Demircan-Tureyen et al. [106]	Fluorescence microscopy	DL	Denoising: CNN	Restoring Fluorescence Microscopy Images	PSNR: 31.37 (+7.21) SSIM: 85.3 % (+39.2 %)
Gour et al. [108]	OCT/OCTA	DL	Denoising: CNN	Generalization of a speckle denoising mode	PSNR: 27.55 (+10) SSIM: 68 % (+60 %)
Liu et al. [113]	Ultrasonography	Math	Contrast enhancement: CLAHE	Gastrointestinal stromal tumors classification	Accuracy: 83.4 % (+3.3 %)
Barata et al. [127]	Dermoscopy	Math	Color normalization: color constancy, SoG	Skin lesion classification	Accuracy: 77.8 % (+14.7 %)

Table 2
Open-access datasets used in multi-centric studies in healthcare.

Dataset	Image modality	Data description	Used by
BRATS 2013	MRI	A brain tumor segmentation dataset (synthetic and real images, n patients = 24). All images are divided into high-grade gliomas (HG) and low-grade gliomas (LG)	[35]
Lung Image Database Consortium (LIDC)	CT	1018 cases of thoracic CT scan with associated manual segmentation of lung lesions	[48]
Public dataset INbreast	Mammography	460 mammography images (115 female, using two views, a mediolateral oblique view, and a craniocaudal view	[51]
CBIS-DDSM dataset	Mammography	2620 mammography images (1925 cases with mass and 695 cases without mass).	[51]
Digital Database for Screening Mammography ((DDSM) dataset	Mammography	2500 studies, with two images of each breast, and patient information	[52]
Mammographic Image Analysis Society (MIAS) dataset	Mammography	20 mammograms obtained from the mediolateral oblique view containing 25 annotated microcalcification clusters	[142]
Nijmegen mammographic databases	Mammography	40 mammograms of both craniocaudal and oblique views from 21 patients	[142]
BCDR-FM dataset (Film Mammography Dataset) from Breast Cancer Digital Repository	Mammography	736 grey-level digitized mammograms (426 benign and 310 malign mass lesions) from 344 patients.	[53]
National Information Society Agency	PET	PET scans (18F-florbetaben or 18F-flute-metamol) and structural T1-weighted 3-dimensional MRI scans of patients with AD or mild cognitive impairment and cognitively normal subjects.	[143]
Alzheimer's Neuroimaging Initiative (ADNI) database	PET	Scans of normal control, mild cognitive impairment (MCI), and AD.	[57]
Mitosis-Atypia	Digital Pathology	11 histology slides with multiple 20x frames per case digitalized with two different scanners	[54,73,76,80,84,144]
MICCAI'16 GlaS challenge [145]	Digital Pathology	colon adenocarcinoma tissue images	[80,84,95]
MICCAI'16 TUPAC challenge [146]	Digital Pathology	500 WSIs of breast cancer patients	[63,81,86]
Camelyon-16 [147]	Digital Pathology	1399 H&E-stained sentinel lymph node sections of breast cancer patients from two different laboratories	[54,63,67,73,83,89,90]
Camelyon-17 [148]	Digital Pathology	1000 WSIs of breast cancer patients from 5 medical centers	[67,73,81,89,90]
SICAPv1 and SICAP-HUVNGR [149]	Digital Pathology	105 H&E WSI of prostate cancer from two hospitals	[73,80,95]
Cancer Genome Atlas (TCGA) [150]	Digital Pathology	publicly funded project with thousands of slides	[65,86,89,90,92,98]
Breast Cancer dataset [151]	Digital Pathology	7909 breast cancer histopathology images acquired on 82 patients	[91]
MoNuSeg dataset [152]	Digital Pathology	30 training images from different tissues with annotated nuclei boundaries	[68,76,82]
BACH dataset [153]	Digital Pathology	400 microscopy images from four different classes of breast cancer	[76]
PAIP2019 [154]	Digital Pathology	100 WSIs of liver cancer patients	[83]
ROSE datasets Retinal OCTA SEgmentation dataset (ROSE)	OCTA	229 OCTA images with vessel annotations at either centerline-level or pixel-level.	[112]
OCTA-500 [155]	OCTA	500 subjects	[112]
The Breast Ultrasound Images Dataset (BUSI)	Ultrasound imaging	780 breast ultrasound images of 600 female patients. The images are categorized into three classes, which are normal, benign, and malignant.	[116]
The BUS dataset	Ultrasound imaging	780 images (normal, benign, and malignant)	[116]
A public dataset	Ultrasound imaging	86 breast cancer ultrasound images	[116]
International Skin Imaging Collaboration (ISIC 2016)	Dermatology imaging	900 annotated dermoscopic images for training (173 melanomas), and 379 images in a held-out test set for evaluation (75 melanomas).	[120,124,134]
International Skin Imaging Collaboration (ISIC) 2017	Dermatology imaging	2000 images with corresponding ground truth images.	[124,130,132,134,135]
International Skin Imaging Collaboration (ISIC) 2019 dataset	Dermatology imaging	25,331 dermoscopy images, 8 classes (i.e., actinic keratosis (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevus (NV), vascular lesion (VASC) and squamous cell carcinoma (SCC)).	[121]
ColorChecker image dataset	Dermatology imaging	568 8-bit sRGB images, most of which have the size 874 × 583.	[156]
EDRA database	Dermatology imaging	482 images, 241 melanomas and 241 benign lesions.	[126,128]
HAM10000	Dermatology imaging	10,015 images of seven distinct types of skin lesions: Actinic Keratoses and Intraepithelial Carcinoma (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevus (NV) and Vascular Lesion (VASC).	[122,124,137]
PH2 dataset	Dermatology imaging	200 dermoscopic images of melanocytic lesions (80 normal nevi, 80 atypical nevi and 40 melanoma).	[124,125,134,135,157]

- Federated learning: federated learning can potentially address the issue of bias from individual centers by allowing models trained on local data to communicate with a central model. This approach enables collaborative learning while preserving data privacy and security [160]. Further research can explore the application of federated learning in multi-centric studies to improve model performance and mitigate biases.
- Emerging imaging modalities: While initial research has explored the use of emerging imaging modalities such as photoacoustic imaging, there is a lack of multi-centric studies in this area. Future research should focus on conducting multi-centric investigations using this modality, which would require the development of data preparation techniques specific to these modalities [161].
- Open-access mindset: despite the significant benefits offered by multi-centric approaches, there is currently a scarcity of open-source multi-centric datasets. Future research should prioritize the creation and sharing of open-source datasets to facilitate technological advancements and enable comparisons among studies working on similar applications. Cultivating an open-access mindset would foster collaboration, and reproducibility, and accelerate progress in the field.

Overall, numerous promising opportunities exist to advance image harmonization techniques and expand their capabilities to new applications in healthcare. Thoughtful innovation in image harmonization techniques, such as federated learning approaches and solutions for emerging modalities, could help enable a more integrated, informative, and transparent analysis of healthcare data across multiple centers and devices.

It is important to acknowledge the limitations of this review. The literature search was restricted to English articles from 2013 to 2023 and did not include a quantitative meta-analysis. Expanding the search criteria and performing statistical comparisons between findings could provide more significant insights into the relative performance of different approaches. However, this review aimed to provide a comprehensive overview of current image harmonization approaches and their impact across different medical imaging applications.

5. Conclusion

Image harmonization enables reliable integration and analysis of diverse imaging data from multiple centers, leading to more robust and generalizable findings that better represent real-world clinical scenarios. By reducing variability and ensuring consistent image characteristics, harmonization techniques improve the reproducibility and comparability of research outcomes, facilitating effective translation of findings into clinical practice. This systematic review aimed to provide a comprehensive overview of image harmonization techniques employed in multi-centric and multi-device studies within the healthcare field. In this work, we identified the most commonly used and effective methods for image harmonization in various imaging modalities, including radiology imaging, nuclear imaging, optical imaging, ultrasound, and dermatology. While current techniques have shown promising results, there is still room for improvement and exploration of new approaches. Future research should focus on refining existing methods, developing standardized protocols, and exploring novel techniques to achieve even greater harmonization and generalizability across diverse datasets. Overall, continued progress in image harmonization methods and their wider adoption can help derive more informative healthcare analytics from integrated multi-source data.

Declaration

All authors have contributed to an acceptable and satisfactory level.

Data availability

Not admissible.

CRediT authorship contribution statement

Silvia Seoni: Writing – review & editing, Visualization, Data curation. **Alen Shahini:** Formal analysis, Data curation. **Kristen M. Meiburger:** Writing – review & editing. **Francesco Marzola:** Writing – original draft. **Giulia Rotunno:** Writing – original draft. **U. Rajendra Acharya:** . **Filippo Molinari:** Writing – review & editing. **Massimo Salvi:** Writing – original draft, Supervision, Conceptualization.

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.

Appendix

Tables A1, A2, A3, A4, A5, A6, A7, A8, A9.

Table A1
Summary of studies ($n = 14$) that apply image harmonization in MR.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Sun et al [34]	11 subjects, different centers	ML	Grayscale normalization: histogram-based	Segmentation	DSC: 69.86 %, +2.35 %
Pereira et al. [35]	BRATS 2013, 2 external test sets, different centers	ML	Grayscale normalization: histogram-based	Segmentation	DSC: 84 %, +6 %; 88 %, +8 %
Ou et al., [36]	Different centers	ML	Grayscale normalization: field of view	Segmentation	Dice: 91 %, +8 %
Jacobsen et al. [37]	EMCC, ISBR, different centers	Math	Grayscale normalization: histogram-based	Segmentation	DSC: 94.54 %, +21.81 %; 87.35 %, +6.69 %
Reiche et al. [42]	CAIN: 27 vol, ADNI: 21 vol, different centers	Math	Denoising: bias field correction	Segmentation	DSC: 91 %; 86.2 %
Carré et al. [30]	TCIA, different centers	Math	Grayscale normalization: z-score	Classification	Accuracy: 82 %, +15 %; 68 %, +2 % AUC: 91 %, +17 %; 72 %, +2 %
Ji et al. [31]	221 subjects, different centers	Math	Grayscale normalization: z-score	Classification	AUC: 88.6 % Accuracy: 81.9 %

(*continued on next page*)

Table A1 (*continued*)

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Modanwal et al. [38]	GE scanner, Siemens scanner	DL	Grayscale normalization: CycleGAN	Segmentation	Dice From GE to Siemens: 98.01 % From Siemens to GE: 98.13 %
Delisle et al. [39]	iSeg, MRBrainS, different centers	DL	Grayscale normalization: GAN	Segmentation	DSC: 87 %, +55.7 %
Koble et al. [40]	iSeg-2017, BRATS 2019, different centers	Math	Grayscale normalization: histogram-based	Segmentation	HE shows better performance than the simple linear transform-based method on both accuracy and DSC
Alnowami et al. [32]	Different scanners, different centers	Math	Grayscale normalization: histogram-based	Classification	Accuracy: 72.10 %, +24.42 %
Foltyn-Dumitru et al. [33]	TCGA: 160 subjects, UCSF: 410 subjects, different centers	Math	Grayscale normalization: z-score	Classification	AUC: 87 %, +42 %; 86 %, +67 %
Albert et al. [41]	6 centers	DL	Grayscale normalization: histogram-based	Segmentation	Dice: 58 %
Ghazvanchahi et al. [43]	ADNI, CAIN, CCNA, different centers	Math	Grayscale normalization: z-score	Segmentation	DSC: 62 %, +2 %

AUC: Area Under the Curve; DL: Deep Learning; DSC: Dice Similarity Coefficient; GAN: Generative Adversarial Network; ML: Machine Learning; n: number of centers.

Table A2
Summary of studies ($n = 4$) that apply image harmonization in CT.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Li et al. [44]	GE: 38 subjects, Philips: 28 subjects, Siemens: 32 subjects	DL	Grayscale normalization: GAN	Classification	AUC: 69 %, +11.1 %
Ligero et al. [45]	43 subjects, different scanners	Math	Resampling: voxel resampling, ComBat	Classification	Mean percentage of robust CT-radiomics features increases from 59.50 % to 89.25 %
Park et al. [47]	79 subjects, 1 external test set	Math	Resampling: voxel resampling, kernel reconstruction	Prediction	AUC: 80.2 %, +3.5 %
Tonneau et al. [48]	4 centers, 1 external test set	Math	Resampling: voxel resampling, intensity clipping (HU), Denoising	Segmentation	AUC: 63 %, +11 %

AUC: Area Under the Curve; DL: Deep Learning; GAN: Generative Adversarial Network; n: number of centers.

Table A3
Summary of studies ($n = 4$) that apply image harmonization in mammography.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Deng et al. [49]	256 images, different centers	ML	Contrast enhancement: fuzzy filtering	Quality metrics	Contrast improvement index: 0.5814, +0.1450
Perre et al. [53]	BCDR, different centers	Math	Contrast enhancement: HE	Classification	AUC: 78.5 %, +2.2 %
Perez et al. [50]	1688 images, $n = 11$	Math	Contrast enhancement: HE	Segmentation	Dice: 74.7 %, +16.1 %
Cao et al. [51]	InBreast: 410 images, 6 centers	Math	Contrast enhancement: CLAHE	Detection	Sensitivity: 91.3 %, +0.9 %
Mechria et al. [52]	DDCM: 600 images, different centers	DL	Grayscale normalization: DCNN	Classification	Accuracy: 92.70 %, +3.47 %

AUC: Area Under the Curve; CLAHE: Contrast Limited Adaptive Histogram Equalization; DCNN: Deep Convolutional Neural Network; DL: Deep Learning; HE: Histogram Equalization; ML: Machine Learning; n: number of centers.

Table A4
Summary of studies ($n = 4$) that apply image harmonization in PET/SPECT imaging.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Thiele et al. [56]	3 scanners, 2 centers	Math	Grayscale normalization: intensity ratio	Classification	Accuracy volumes: 85 %, +6 %; 88 %, +8 %; Accuracy scans: 86 %, +6 %; 86 %, +8 %
Lee et al. [57]	ADNI, 251 subjects, 1 centers	Math	Grayscale normalization: intensity ratio	Classification	Accuracy: 87 %
Ren et al. [58]	1 centers	Math	Grayscale normalization: standardized uptake values	Segmentation	DSC: 77.87 %, +1.16 %
Kang et al. [143]	148 images, 3 centers	DL	Grayscale normalization: DNN	Correlation	ICC: 99.2 %; 98.9 %; 98.5 %

DL: Deep Learning; DNN: Deep Neural Network; DSC: Dice Similarity Coefficient; ICC: Intraclass Correlation Coefficient; ML: Machine Learning; n: number of centers.

Table A5
Summary of studies ($n = 38$) that apply image harmonization in digital pathology.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Khan et al. [75]	60 WSI, 3 centers	ML	Color normalization: color deconvolution	Classification, Segmentation	Accuracy: 96 %, +12 % Dice: 80 %, +2 %
Tam et al. [62]	434 patients, different centers	Math	Color normalization: histogram-based	Classification	Accuracy: 66.2 %, +14.8 %

(*continued on next page*)

Table A5 (*continued*)

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Bejnordi et al. [74]	155 WSI, 6 centers	ML	Color normalization: color deconvolution	Classification	AUC: 94.4 %, +63.4 %
Bentaieb et al. [84]	55 WSI, different centers	DL	Color normalization: color deconvolution	Classification	Accuracy: 61.7 %, +7 %
Alsubaie et al. [66]	20 WSI, different centers	Math	Color normalization: color deconvolution	Detection	F1-score: 41.6 %, no improvement
Janowczyk et al. [78]	Different centers	DL	Color normalization: AutoEncoder	Segmentation	Dice: 76.5 %, +14.5 %
Shaban et al. [88]	1 center	DL	Color normalization: CycleGAN	Classification	AUC: 79 %, +35 %
Anghel et al. [63]	1 center	Math	Color normalization: color deconvolution	Classification	F1: 79 %, +5 %
Zheng et al. [67]	130 WSI	Math	Color normalization: color deconvolution	Classification	AUC: 91.4 %, +7.2 %
Otalora et al. [98]	TMA-Zürich: 50 WSI, 6 centers	DL	Color normalization: GAN	Detection, Classification	F1: 82.4 %, +11 %; AUC: 91.5 %, +11 % F1: 62.22 %, +14 %; AUC: 69.1 %, +5 %
Lafarge et al. [86]	20 centers	DL	Color normalization: GAN	Detection, Segmentation	F1: 60 %, +48 % F1: 85 %, +2 %
Shrivastava et al. [85]	16 WSI, 2 centers	DL	Color normalization: GAN	Quality metrics	SSIM: 99.6 %, better performance on external test set
Zaneta Swiderska-Chadaj et al. [79]	135 WSI, 2 centers	DL	Color normalization: CycleGAN	Classification	Accuracy: 93 %, +10 %; 92 %, +7 %
Shafei et al. [76]	BACH Dataset	ML	Color normalization: Spatial Finite Mixture of Skew-Normal distributions	Classification	Accuracy: 79.75 %, +7 %
Salvi et al. [64]	Camelyon-16	Math	Color normalization: color deconvolution	Detection	Accuracy: 92.87 %, +11 %
Tellez et al. [81]	200 WSI, 6 centers	DL	Color normalization: GAN	Detection, Classification	Color normalization is crucial to obtaining top classification performance on external test sets.
Salehi et al. [87]	16 WSI, 2 centers	DL	Color normalization: Pix2Pix	Quality metrics	SSIM: 84.5 %, +4 %
Perez et al. [73]	2 centers	DL	Color normalization: GAN	Segmentation	Dice: 71.75 %, +13 %; 61.78 %, +7 %
Cong et al. [89]	TCGA: 2310 patches	DL	Color normalization: Pix2Pix	Classification	AUC: 96.7 %, +3 %
Kang et al. [54]	Camelyon-16: 170 WSI, $n = 1$	DL	Color normalization: CycleGAN	Classification	AUC: 89.5 %, +21 %
Marini et al. [80]	7 datasets, 7+ centers	DL	Color normalization: CNN	Classification	Cohen's k-score: 0.532, +0.108; 0.474, +0.422
Faryna et al. [93]	Camelyon-17, 2 centers	Math	Color normalization: augmentation transforms	Classification	AUC: 96.4 %, +51.4 %
Perez-Bueno et al. [162]	270 WSI, 2 centers	ML	Color normalization: Gaussian priors, Bayesian inference	Classification	AUC: 96.56 %, +1.6 %
Mahmood et al. [65]	TCGA	Math	Color normalization: color deconvolution	Segmentation	Dice: 80.84 %, +3.4 % F1: 85.47 %, +3.8 %
Boschman et al. [91]	113 WSI, different centers	DL	Color normalization: color deconvolution, SNMF, GAN	Classification	AUC: 94 %, +25 %; 98 %, +20 %
Jeong et al. [83]	PAIP19, Camelyon-16	ML	Color normalization: SNMF	Quality metrics	MS-SSIM: 96.97 %; 95.78 % PSNR: 23.85; 19.79 PCC: 94.85 %; 95.68 % Better performance on external test set
Cong et al. [90],	Camelyon-17	DL	Color normalization: GAN	Quality metrics Classification	Performance improvement by 5 % to 10 % on external test sets
Bouteldja et al. [99]	5 WSI, 3 centers	DL	Color normalization: Color augmentation, CycleGAN	Segmentation	Dice: 83.2 %, +0.6 %; 86.5 %, +2.3 %; 82 %, +1.3 %
Altini et al. [92]	10 WSI, 1 center	DL	Color normalization: SNMF, GAN, CycleGAN	Classification	Accuracy: 82.05 %, +1.8 %; 83.70 %, +2.65 %; 78.08 %, +6.4 %
Marini et al. [95]	12 datasets, 12+ centers	Math	Color normalization: color augmentation	Classification	Cohen's k-score: 0.432, +0.165; 0.477, +0.267
Martos et al. [68]	TCGA: 16 WSI	Math	Color normalization: color deconvolution	Detection	F1-score: 90.7 %, +7.1 %
Sun et al. [82]	6 WSI	Math	Color normalization: z-score	Segmentation	Dice: 82.2 %, +2.4 % F1: 86.8 %, +3.6 %
Tolkach et al. 2023 [100]	4 centers	DL	Color normalization	Detection	AUROC: +0.12
Dammaka et al. 2023 [96]	35 centers	Math	Color normalization	Classification	AUC: +0.20 (from 0.7 to 0.9)
Wang et al. 2023 [69]	4 external datasets	Math	Color normalization	Detection	
Huang et al. 2023 [97]	2 centers	Math	Color normalization	Classification / Segmentation	Performance: +8.3 % and 15.3 %
Alhassan et al. 2023 [70]	4 centers	Math	Color normalization	Segmentation	Accuracy: +6.5 %
Faryna et al. 2023 [94]	25 centers	Math	Color normalization	Detection	AUC: +50 %
Bazargani et al. 2023 [71]	2 centers	Math	Color normalization	Classification	AUC:+ 0.03 and 0.05 in internal and external dataset
Marini et al. 2023 [95]		DL	Color normalization	Classification	Classification performance: 26 %

AUC: Area Under the Curve; CNN: Convolutional Neural Network; DL: Deep Learning; GAN: Generative Adversarial Network; ML: Machine Learning; MS-SSIM: Multi Scale Structural Similarity Index Measure; n: number of centers; PCC: Pearson Correlation Coefficient; PSNR: Peak Signal-to-Noise Ratio; SNMF: Sparse Non-negative Matrix Factorization; SSIM: Structural Similarity Index Measure; WSI: Whole Slide Image.

Table A6
Summary of studies ($n = 6$) that apply image harmonization in fluorescence microscopy.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Yang et al. [101]	140 images, 2 centers	DL	Denoising: teacher-based DL model	Quality metrics	PSNR: 34.02, +7.69
Zhang et al. [105]	FMD, 2 centers	DL	Denoising: CNN (Noise2Noise)	Quality metrics	PSNR: 33.02, +10.31; 36.35, +5.68 SSIM: 91.09 %, +46.68 %; 94.41 %, +15.39 %
Broaddus et al. [104]	100 images, 2 centers	DL	Denoising: CNN (Noise2Void)	Quality metrics	PSNR: 29.73 SSIM: 91.3 %
Yang et al. [103]	RxRx1	Math	Grayscale normalization: z-score	Classification	Accuracy: 74.58 %
Mannam et al. [102]	Widefield2SIM: 360 images	DL	Denoising: CNN, (Noise2Noise)	Quality metrics	PSNR: +8.25
Demircan-Tureyen et al. [106]	FMD, 7 centers	DL	Denoising: CNN	Quality metrics	PSNR: 31.37, +7.21 SSIM: 85.3 %, +39.2 %

CNN: Convolutional Neural Network; DL: Deep Learning; ML: Machine Learning; n: number of centers; PSNR: Peak Signal-to-Noise Ratio; SSIM: Structural Similarity Index Measure.

Table A7
Summary of studies ($n = 6$) that apply image harmonization in OCT/OCTA.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Venhuizen et al. [110]	30 vol, 4 scanners	Math	Resampling: pixel resampling	Segmentation	Dice: 75.4 %
Shi et al. [107]	Topcon: 11 vol, 4 scanners	DL	Denoising: CNN	Quality metrics	PSNR: 40.17, +14 CNR: 9.67, +5.3
Bogunovic et al. [111]	RETOUCH: 42 vol, 3 scanners: Cirrus, Spectralis, Topcon)	Math	Grayscale normalization: histogram-based	Segmentation	DSC: 82 %; 75 %; 74 %
Romo-Bucheli et al. [109]	2 scanners (Cirrus, Spectralis)	DL	Grayscale normalization: CycleGAN	Segmentation	Dice Cirrus as Test Set: 48 %, +47 %; 55 %, +54 %; 85 %, +29 % Spectralis as Test Set: 59 %, no improvement; 66 %, +11 %; 88 %, +4 %
Gour et al. [108]	BSDS500 and Topcon dataset	DL	Denoising: CNN	Quality metrics	PSNR: 27.55, +10 SSIM: 68 %, +60 %
Ma et al. [112]	OCTA-500, ROSE dataset, different scanners	Math	Contrast enhancement: CLAHE	Segmentation	Dice: 76.04 %, +0.4 %

CLAHE: Contrast Limited Adaptive Histogram Equalization; CNN: Convolutional Neural Network; CNR: Contrast-to-Noise Ratio; DL: Deep Learning; DSC: Dice Similarity Coefficient; GAN: Generative Adversarial Network; PSNR: Peak Signal-to-Noise Ratio; SSIM: Structural Similarity Index Measure.

Table A8
Summary of studies ($n = 5$) that apply image harmonization in US imaging.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Liu et al. [113]	914 subjects, 18 centers	Math	Contrast enhancement: CLAHE	Classification	Accuracy: 83.4 %, +3.3 %
Ren et al. [114]	1370 subjects, 3 centers (1 external)	Math	Grayscale normalization: z-score	Segmentation	Accuracy: 81.82 %
Homayoun et al. [115]	1259 subjects, 3 centers (2 external)	Math	Grayscale normalization: 3-sigma	Classification	Accuracy: 88.8 %; 87.9 %
Du et al. [117]	175 subjects, 2 centers	Math	Grayscale normalization: z-score	Segmentation	n/a
Sirjani et al. [116]	1656 subjects, 5 centers (1 external)	Math	Grayscale normalization: min-max scaling	Classification	Accuracy: 91 %

CLAHE: Contrast Limited Adaptive Histogram Equalization; n: number of centers.

Table A9
Summary of studies ($n = 19$) that apply image harmonization in dermoscopy.

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
Abbas et al. [129]	EDRA: 100 images, 3 centers	Math	Color normalization: color constancy	Detection	Recall: 94.25 %, +10 %
Barata et al. [127]	EDRA: 482 images, 3 centers	Math	Color normalization: color constancy, SoG	Classification	Accuracy: 77.8 %, +14.7 %
Barata et al. [128]	EDRA: 482 images, 3 centers	Math	Color normalization: color constancy, SoG	Classification	Accuracy: 73.4 %, +14.6 %
Codella et al. [118]	ISIC 2016: 379 images, different centers	Math	Resampling: resize and cropping	Classification	Accuracy: 76 %
Galdran et al. [138]	ISIC 2017: 500 images, different centers	Math	Color normalization: color constancy augmentation	Segmentation	Accuracy: 94.8 % Dice: 84.6 %
Yu et al. [120]	ISIC 2017: 379 images, different centers	Math	Color normalization: z-score	Classification	Accuracy: 82.97 %
Olga et al. [131]	ISIC 2016: 379 images, different centers	Math	Color normalization: color constancy	Segmentation	Dice: 41 %, +2 %
Ng et al. [130]	ISIC 2017: 600 images, different centers	Math	Color normalization: color constancy, SLRMSR	Segmentation	Accuracy: 93.37 %, +0.24 %

(*continued on next page*)

Table A9 (*continued*)

Author, year	Dataset	Technique	Strategy	Task	Results/Impact of normalization
					Dice: 84.41 %, +2.21 %
Yuan et al. [134]	ISIC 2017: 600 images, different centers	Math	Color normalization: color constancy	Segmentation	Accuracy: 93.4 %, +0.3 % Dice: 84.9 %, +0.9 %
Zhang et al. [132]	ISIC 2017: 600 images, different centers	Math	Color normalization: color constancy, SoG	Segmentation	Accuracy: 91.8 %, +0.1 %
Goyal et al. [135]	ISIC 2017: 600 images, PH2: 200 images, different centers	Math	Color normalization: color constancy	Segmentation	Accuracy: 94.08 %; 93.8 % Dice: 87.14 %, 90.7 %
Gong et al. [121]	ISIC 2019, 5066 images, different centers	Math	Color normalization: z-score	Classification	Accuracy: 97.5 %
Zafar et al. [125]	ISIC 2017: 600 images, PH2: 200 images, different centers	Math	Color normalization: z-score	Segmentation	Dice: 85.8 %; 92.4 %
Shahin Ali et al. [122]	HAM10000: 1000 images, different centers	Math	Color normalization: z-score	Classification	Accuracy: 91.43 %
Xin et al. [123]	HAM10000, different centers	Math	Color normalization: z-score	Classification	Accuracy: 94.3 %
Salvi et al. [137]	HAM10000: 8715 images, different centers	DL	Color normalization: GAN	Classification, Segmentation	Accuracy: 79.2 %, +2.6 % Dice: 90.9 %, +8.8 %
Azad et al. [119]	ISIC 2018: 520 images, different centers	DL	Resampling: cropping	Segmentation	Accuracy: 96.5 %, +0.7 % Dice: 91.31 %, +1.8 %
Behara et al. [126]	ISIC 2017 dataset: 2000 images	Math	Color normalization	Classification	Accuracy: 99.38 %
Gajera et al. [124]	PH2: 60 images, ISIC 2016: 379 images, ISIC 2017: 600 images, HAM10000: 3000 images, different centers	Math	Color normalization: z-score	Classification, Segmentation	Accuracy: 98.33 % Dice: 96 %

DL: Deep Learning; GAN: Generative Adversarial Network; n: number of centers; SLRMSR: Smart Light Random Memory Spray Retinex; SoG: Shades of Gray.

References

[1] A. Visvizi, O. Troisi, M. Grimaldi, Big data and decision-making: how big data is relevant across fields and domains, Big Data Decisi.-Mak.: Appl. Uses Public Private Sect. (2023) 1–11, https://doi.org/10.1108/978-1-80382-551-920231001/FULL/XML. Jan.

[2] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005. Elsevier B.V.Dec. 01.

[3] M. Salvi, et al., Histopathological classification of canine cutaneous round cell tumors using deep learning: a multi-center study, Front. Vet. Sci. 8 (2021), https://doi.org/10.3389/fvets.2021.640944. Mar.

[4] J. Yao, et al., A multi-center milestone study of clinical vertebral CT segmentation, Comput. Med. Imag. Graph. 49 (2016) 16–28, https://doi.org/10.1016/j.compmedimag.2015.12.006. Apr.

[5] N. Michielli, et al., Stain normalization in digital pathology: clinical multi-center evaluation of image quality, J. Pathol. Inform. 13 (2022), https://doi.org/10.1016/j.jpi.2022.100145. Jan.

[6] Z. Liu, F. Wu, Y. Wang, M. Yang, X. Pan, FedCL: federated contrastive learning for multi-center medical image classification, Pattern. Recognit. 143 (2023), https://doi.org/10.1016/j.patcog.2023.109739. Nov.

[7] J. Xu, et al., Deep reconstruction-recoding network for unsupervised domain adaptation and multi-center generalization in colonoscopy polyp detection, Comput. Method. Program. Biomed. 214 (2022), https://doi.org/10.1016/j.cmpb.2021.106576. Feb.

[8] K.D. Kim, et al., Enhancing deep learning based classifiers with inpainting anatomical side markers (L/R markers) for multi-center trials, Comput. Method. Program. Biomed. 220 (2022), https://doi.org/10.1016/j.cmpb.2022.106705. Jun.

[9] M. Salvi, U.R. Acharya, F. Molinari, K.M. Meiburger, The impact of pre- and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis, Comput. Biol. Med. 128 (2021), https://doi.org/10.1016/j.compbiomed.2020.104129. Elsevier LtdJan. 01.

[10] S.A. Mali, et al., Making radiomics more reproducible across scanner and imaging protocol variations: a review of harmonization methods, J. Pers. Med. 11 (9) (2021), https://doi.org/10.3390/jpm11090842. MDPISep. 01.

[11] L.G. Nyú, J.K. Udupa, On standardizing the MR image intensity scale, Magn. Reson. Med. 42 (6) (1999) 1072–1081, https://doi.org/10.1002/(SICI)1522-2594(199912)42:6<1072::AID-MRM11>3.0.CO;2-M.

[12] P. Vasuki J. Kanimozhi M. Balkis Devi KLNCIT, S. Klncit, S. Tamilnadu, and I. Tamilnadu, "A survey on image preprocessing techniques for diverse fields of medical imagery".

[13] A. Makandar, B. Halalli, and R. Scholar, "A review on preprocessing techniques for digital mammography images," 2015.

[14] T.A. Azevedo Tosta, P.R. de Faria, L.A. Neves, M.Z. do Nascimento, Computational normalization of H&E-stained histological images: progress, challenges and future potential, Artif. Intell. Med. 95 (2019) 118–132, https://doi.org/10.1016/j.artmed.2018.10.004. Elsevier B.V.Apr. 01.

[15] M.S. Pinto, et al., Harmonization of brain diffusion MRI: concepts and methods, Front Neurosci 14 (2020), https://doi.org/10.3389/fnins.2020.00396. Frontiers Media S.A.May 06.

[16] L. Saba, et al., The present and future of deep learning in radiology, Eur. J. Radiol. 114 (2019) 14–24, https://doi.org/10.1016/J.EJRAD.2019.02.038. May.

[17] K.G. van Leeuwen, M. de Rooij, S. Schalekamp, B. van Ginneken, M.J.C. M. Rutten, How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr. Radiol. 52 (11) (2022) 2087–2093, https://doi.org/10.1007/S00247-021-05114-8/FIGURES/2. Oct.

[18] R. Seifert, M. Weber, E. Kocakavuk, C. Rischpler, D. Kersting, Artificial intelligence and machine learning in nuclear medicine: future perspectives, Semin. Nucl. Med. 51 (2) (2021) 170–177, https://doi.org/10.1053/J.SEMNUCLMED.2020.08.003. Mar.

[19] C. Bruce, et al., Transforming diagnostics: the implementation of digital pathology in clinical laboratories, Histopathology (2024), https://doi.org/10.1111/HIS.15178.

[20] K.A. Heger, S.M. Waldstein, Artificial intelligence in retinal imaging: current status and future prospects, Expert. Rev. Med. Devices 21 (1–2) (2024) 73–89, https://doi.org/10.1080/17434440.2023.2294364. Feb.

[21] J.F. Zhang, et al., The application of optical coherence tomography angiography in cerebral small vessel disease, ischemic stroke, and dementia: a systematic review, Front. Neurol. 11 (2020) 560038, https://doi.org/10.3389/FNEUR.2020.01009/BIBTEX. Sep.

[22] P.D. Barua, et al., Multilevel deep feature generation framework for automated detection of retinal abnormalities using OCT images, Entropy 23 (12) (2021) 1651, https://doi.org/10.3390/E23121651. *2021, Vol. 23, Page 1651*Dec.

[23] S. Arslan, et al., Attention TurkerNeXt: investigations into bipolar disorder detection using OCT images, Diagnostics 13 (22) (2023), https://doi.org/10.3390/diagnostics13223422. Nov.

[24] K.M. Meiburger, M. Salvi, G. Rotunno, W. Drexler, M. Liu, Automatic segmentation and classification methods using optical coherence tomography angiography (Octa): a review and handbook, Appl. Sci. (Switzerland) 11 (20) (2021), https://doi.org/10.3390/app11209734. MDPIOct. 01.

[25] Y.T. Shen, L. Chen, W.W. Yue, H.X. Xu, Artificial intelligence in ultrasound, Eur. J. Radiol. 139 (2021) 109717, https://doi.org/10.1016/J.EJRAD.2021.109717. Jun.

[26] E. Kaplan, et al., PFP-LHCINCA: pyramidal fixed-size patch-based feature extraction and Chi-square iterative neighborhood component analysis for automated fetal sex classification on ultrasound images, Hindawi Contrast Media Mol. Imag. 2022 (2022), https://doi.org/10.1155/2022/6034971.

[27] E. Kaplan, et al., Automated BI-RADS classification of lesions using pyramid triple deep feature generator technique on breast ultrasound images, Med. Eng. Phys. 108 (2022) 1350–4533, https://doi.org/10.1016/j.medengphy.2022.103895.

[28] K. Liopyris, S. Gregoriou, J. Dias, A.J. Stratigos, Artificial intelligence in dermatology: challenges and perspectives, Dermatol. Ther. (Heidelb) 12 (12) (2022) 2637–2651, https://doi.org/10.1007/S13555-022-00833-8/FIGURES/3. Dec.

[29] V.V. Pai, R.B. Pai, V.V. Pai, R.B. Pai, Artificial intelligence in dermatology and healthcare: an overview, Indian J. Dermatol. Venereol. Leprol. 87 (4) (2021) 457–467, https://doi.org/10.25259/IJDVL_518_19. Jun.
[30] A. Carré, et al., Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics, Sci. Rep. 10 (1) (2020), https://doi.org/10.1038/s41598-020-69298-z. Dec.
[31] X. Ji, et al., Bi-parametric magnetic resonance imaging based radiomics for the identification of benign and malignant prostate lesions: cross-vendor validation, Phys. Eng. Sci. Med. 44 (3) (2021) 745–754, https://doi.org/10.1007/s13246-021-01022-1. Sep.
[32] M. Alnowami, E. Taha, S. Alsebaeai, S.Muhammad Anwar, A. Alhawsawi, MR image normalization dilemma and the accuracy of brain tumor classification model, J. Radiat. Res. Appl. Sci. 15 (3) (2022) 33–39, https://doi.org/10.1016/j.jrras.2022.05.014. Sep.
[33] M. Foltyn-Dumitru, et al., Impact of signal intensity normalization of MRI on the generalizability of radiomic-based prediction of molecular glioma subtypes, Eur. Radiol. (2023), https://doi.org/10.1007/s00330-023-10034-2.
[34] X. Sun, et al., Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions, Biomed. Eng. Online 14 (1) (2015), https://doi.org/10.1186/s12938-015-0064-y. Jul.
[35] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEe Trans. Med. ImAging 35 (5) (2016) 1240–1251, https://doi.org/10.1109/TMI.2016.2538465. May.
[36] Y. Ou, et al., Field of view normalization in multi-site brain MRI, Neuroinformatics 16 (3–4) (2018) 431–444, https://doi.org/10.1007/s12021-018-9359-z. Oct.
[37] N. Jacobsen, A. Deistung, D. Timmann, S.L. Goericke, J.R. Reichenbach, D. Güllmar, Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network, Z. Med. Phys. 29 (2) (2019) 128–138, https://doi.org/10.1016/j.zemedi.2018.11.004. May.
[38] G. Modanwal, A. Vellal, M.A. Mazurowski, Normalization of breast MRIs using cycle-consistent generative adversarial networks, Comput. Methods Programs Biomed. 208 (2021), https://doi.org/10.1016/j.cmpb.2021.106225. Sep.
[39] P.L. Delisle, B. Anctil-Robitaille, C. Desrosiers, H. Lombaert, Realistic image normalization for multi-domain segmentation, Med. Image Anal. 74 (2021), https://doi.org/10.1016/j.media.2021.102191. Dec.
[40] A. Koble, et al., Identifying the most suitable histogram normalization technique for machine learning based segmentation of multispectral brain MRI data, in: IEEE AFRICON Conference, Institute of Electrical and Electronics Engineers Inc., 2021, https://doi.org/10.1109/AFRICON51333.2021.9570990. Sep.
[41] S. Albert, et al., Comparison of image normalization methods for multi-site deep learning, Appl. Sci. (Switzerland) 13 (15) (2023), https://doi.org/10.3390/app13158923. Aug.
[42] B. Reiche, A.R. Moody, A. Khademi, Pathology-preserving intensity standardization framework for multi-institutional FLAIR MRI datasets, Magn. Reson. ImAging 62 (2019) 59–69, https://doi.org/10.1016/j.mri.2019.05.001. Oct.
[43] A. Ghazvanchahi, P.J. Maralani, A.R. Moody, and A. Khademi, "Effect of intensity standardization on deep learning for WML segmentation in Multi-centre FLAIR MRI," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.03827.
[44] Y. Li, et al., Normalization of multicenter CT radiomics by a generative adversarial network method, Phys. Med. Biol. 66 (5) (2021), https://doi.org/10.1088/1361-6560/ab8319. Mar.
[45] M. Ligero, et al., Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis, Eur. Radiol. 31 (3) (2021) 1460–1470, https://doi.org/10.1007/s00330-020-07174-0. Mar.
[46] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics. 8 (1) (2007) 118–127, https://doi.org/10.1093/biostatistics/kxj037. Jan.
[47] D. Park, et al., Importance of CT image normalization in radiomics analysis: prediction of 3-year recurrence-free survival in non-small cell lung cancer, Eur. Radiol. 32 (12) (2022) 8716–8725, https://doi.org/10.1007/s00330-022-08869-2. Dec.
[48] M. Tonneau, et al., Generalization optimizing machine learning to improve CT scan radiomics and assess immune checkpoint inhibitors' response in non-small cell lung cancer: a multicenter cohort study, Front. Oncol. 13 (2023), https://doi.org/10.3389/fonc.2023.1196414.
[49] H. Deng, W. Deng, X. Sun, M. Liu, C. Ye, X. Zhou, Mammogram enhancement using intuitionistic fuzzy sets, IEEE Trans. Biomed. Eng. 64 (8) (2017) 1803–1814, https://doi.org/10.1109/TBME.2016.2624306. Aug.
[50] F.J. Pérez-Benito, et al., A deep learning system to obtain the optimal parameters for a threshold-based breast and dense tissue segmentation, Comput. Methods Programs Biomed. 195 (2020), https://doi.org/10.1016/j.cmpb.2020.105668. Oct.
[51] H. Cao, S. Pu, W. Tan, J. Tong, Breast mass detection in digital mammography based on anchor-free architecture, Comput. Methods Programs Biomed. 205 (2021), https://doi.org/10.1016/j.cmpb.2021.106033. Jun.
[52] H. Mechria, K. Hassine, M.S. Gouider, Effect of Denoising on Performance of Deep Convolutional Neural Network For Mammogram Images Classification, in: Procedia Computer Science, Elsevier B.V., 2022, pp. 2345–2352, https://doi.org/10.1016/j.procs.2022.09.293.
[53] A.C. Perre, L. Alexandre, L. Freire, The Influence of Image Normalization in Mammographic Classification with CNNs, in: 23rd Portuguese Conference on Pattern Recognition, RECPAD 2017, 2017. Accessed: Jan. 24, 2024. [Online]. Available: https://ubibliorum.ubi.pt/handle/10400.6/8237.
[54] H. Kang, et al., StainNet: a fast and robust stain normalization network, Front. Med. (Lausanne) 8 (2021), https://doi.org/10.3389/fmed.2021.746307. Nov.
[55] B. Fischl, FreeSurfer, Neuroimage 62 (2) (2012) 774–781, https://doi.org/10.1016/j.neuroimage.2012.01.021. Aug. 15.
[56] F. Thiele, S. Young, R. Buchert, F. Wenzel, Voxel-based classification of FDG PET in dementia using inter-scanner normalization, Neuroimage 77 (2013) 62–69, https://doi.org/10.1016/j.neuroimage.2013.03.031. Aug.
[57] S.Y. Lee, H. Kang, J.H. Jeong, D.Y. Kang, Performance evaluation in [18F] Florbetaben brain PET images classification using 3D Convolutional Neural Network, PLoS ONE 16 (10 October 2021) (2021), https://doi.org/10.1371/journal.pone.0258214. Oct.
[58] J. Ren, B.N. Huynh, A.R. Groendahl, O. Tomic, C.M. Futsaether, S.S. Korreman, PET normalizations to improve deep learning auto-segmentation of head and neck tumors in 3D PET/CT, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2022, pp. 83–91, https://doi.org/10.1007/978-3-030-98253-9_7.
[59] M. Salvi, et al., Impact of stain normalization on pathologist assessment of prostate cancer: a comparative study, Cancers. (Basel) 15 (5) (2023), https://doi.org/10.3390/cancers15051503. Mar.
[60] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley, "Color Transfer between Images".
[61] Marc Macenko, et al., A Method for Normalizing Histology Slides for Quantitative Analysis, IEEE, 2009.
[62] A. Tam, J. Barker, D. Rubin, A method for normalizing pathology images to improve feature extraction for quantitative pathology, Med. Phys. 43 (1) (2016) 528–537, https://doi.org/10.1118/1.4939130. Jan.
[63] A. Anghel, et al., A high-performance system for robust stain normalization of whole-slide images in histopathology, Front. Med. (Lausanne) 6 (2019), https://doi.org/10.3389/fmed.2019.00193. Sep.
[64] M. Salvi, N. Michielli, F. Molinari, Stain Color Adaptive Normalization (SCAN) algorithm: separation and standardization of histological stains in digital pathology, Comput. Methods Programs Biomed. 193 (2020), https://doi.org/10.1016/j.cmpb.2020.105506. Sep.
[65] T. Mahmood, et al., Accurate segmentation of nuclear regions with multi-organ histopathology images using artificial intelligence for cancer diagnosis in personalized medicine, J. Pers. Med. 11 (6) (2021), https://doi.org/10.3390/jpm11060515. Jun.
[66] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, N.M. Rajpoot, Stain deconvolution using statistical analysis of multi-resolution stain colour representation, PLoS ONE 12 (1) (2017), https://doi.org/10.1371/journal.pone.0169875. Jan.
[67] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, C. Xue, Adaptive color deconvolution for histological WSI normalization, Comput. Methods Programs Biomed. 170 (2019) 107–120, https://doi.org/10.1016/j.cmpb.2019.01.008. Mar.
[68] O. Martos, et al., Optimized detection and segmentation of nuclei in gastric cancer images using stain normalization and blurred artifact removal, Pathol. Res. Pract. 248 (2023), https://doi.org/10.1016/j.prp.2023.154694. Aug.
[69] X. Wang, et al., A generalizable and robust deep learning algorithm for mitosis detection in multicenter breast histopathological images, Med. Image Anal. 84 (2023), https://doi.org/10.1016/j.media.2022.102703. Feb.
[70] A.M. Alhassan, Driving training-based optimization-multitask fuzzy C-means (DTBO-MFCM) image segmentation and robust deep learning algorithm for multicenter breast histopathological images, IEEe Access. 11 (2023) 136350–136360, https://doi.org/10.1109/ACCESS.2023.3335667.
[71] R. Bazargani, et al., A novel H and E color augmentation for domain invariance classification of unannotated histopathology prostate cancer images, in: SPIE-Intl Soc Optical Eng, 2023, p. 35, https://doi.org/10.1117/12.2654040. Apr.
[72] M. Gavrilovic, et al., Blind color decomposition of histological images, IEEe Trans. Med. Imaging 32 (6) (2013) 983–994, https://doi.org/10.1109/TMI.2013.2239655.
[73] J.C. Gutiérrez Pérez, D. Otero Baguer, P. Maass, StainCUT: stain normalization with contrastive learning, J. ImAging 8 (7) (2022), https://doi.org/10.3390/jimaging8070202. Jul.
[74] B.E. Bejnordi, et al., Stain specific standardization of whole-slide histopathological images, IEEe Trans. Med. ImAging 35 (2) (2016) 404–415, https://doi.org/10.1109/TMI.2015.2476509. Feb.
[75] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, IEEE Trans. Biomed. Eng. 61 (6) (2014) 1729–1738, https://doi.org/10.1109/TBME.2014.2303294.
[76] S. Shafiei, A. Safarpoor, A. Jamalizadeh, H.R. Tizhoosh, Class-agnostic weighted normalization of staining in histopathology images using a spatially constrained mixture model, IEEe Trans. Med. ImAging 39 (11) (2020) 3355–3366, https://doi.org/10.1109/TMI.2020.2992108. Nov.
[77] Farhad Ghazvinian Zanjani, Svitlana Zinger, Babak Ehteshami Bejnordi, Jeroen AWM van der Laak, Pter H.N. de With, Stain normalization of histopathology images using generative adversarial networks, IEEE, 2018.
[78] A. Janowczyk, A. Basavanhally, A. Madabhushi, Stain Normalization using Sparse AutoEncoders (StaNoSA): application to digital pathology, Comput. Med. Imag. Graphic. 57 (2017) 50–61, https://doi.org/10.1016/j.compmedimag.2016.05.003. Apr.
[79] Z. Swiderska-Chadaj, et al., Impact of rescanning and normalization on convolutional neural network performance in multi-center, whole-slide classification of prostate cancer, Sci. Rep. 10 (1) (2020), https://doi.org/10.1038/s41598-020-71420-0. Dec.

[80] N.O. Marini, M. Atzori, S. Otálora, S. Marchand-Maillet, and H. Müller, "H&E-adversarial network: a convolutional neural network to learn stain-invariant features through Hematoxylin & Eosin regression".
[81] D. Tellez, et al., Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, Med. Image Anal. 58 (2019), https://doi.org/10.1016/j.media.2019.101544. Dec.
[82] M. Sun, W. Zou, Z. Wang, S. Wang, Z. Sun, An automated framework for histopathological nucleus segmentation with deep attention integrated networks, IEEE/ACM. Trans. Comput. Biol. Bioinform. (2022), https://doi.org/10.1109/TCBB.2022.3233400.
[83] J. Jeong, K.D. Kim, Y. Nam, C.E. Cho, H. Go, N. Kim, Stain normalization using score-based diffusion model through stain separation and overlapped moving window patch strategies, Comput. Biol. Med. 152 (2023), https://doi.org/10.1016/j.compbiomed.2022.106335. Jan.
[84] A. BenTaieb, G. Hamarneh, Adversarial stain transfer for histopathology image analysis, IEEe Trans. Med. ImAging 37 (3) (2018) 792–802, https://doi.org/10.1109/TMI.2017.2781228. Mar.
[85] A. Shrivastava et al., "Self-attentive adversarial stain normalization," Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.01963.
[86] M.W. Lafarge, J.P.W. Pluim, K.A.J. Eppenhof, M. Veta, Learning domain-invariant representations of histological images, Front. Med. (Lausanne) 6 (2019), https://doi.org/10.3389/fmed.2019.00162. Jul.
[87] P. Salehi and A. Chalechale, "Pix2Pix-based Stain-to-Stain translation: a solution for robust stain normalization in histopathology images analysis." [Online]. Available: https://github.com/pegahsalehi/Stain-to-Stain-Translation.
[88] M.Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni, *Staingan: stain style transfer for digital histological images*.
[89] C. Cong, S. Liu, A. Di Ieva, M. Pagnucco, S. Berkovsky, Y. Song, Texture enhanced generative adversarial network for stain normalisation in histopathology images, in: Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society, 2021, pp. 1949–1952, https://doi.org/10.1109/ISBI48211.2021.9433860. Apr.
[90] C. Cong, S. Liu, A. Di Ieva, M. Pagnucco, S. Berkovsky, Y. Song, Colour adaptive generative networks for stain normalisation of histopathology images, Med. Image Anal. 82 (2022), https://doi.org/10.1016/j.media.2022.102580. Nov.
[91] J. Boschman, et al., The utility of color normalization for AI-based diagnosis of hematoxylin and eosin-stained pathology images, J. Pathol. 256 (1) (2022) 15–24, https://doi.org/10.1002/path.5797. Jan.
[92] N. Altini, et al., The role of unpaired image-to-image translation for stain color normalization in colorectal cancer histology classification, Comput. Method. Program. Biomed. 234 (2023), https://doi.org/10.1016/j.cmpb.2023.107511. Jun.
[93] K. Faryna, J. Van Der Laak, and G. Litjens, "Tailoring automated data augmentation to H&E-stained histopathology." [Online]. Available: https://github.com/DIAGNijmegen/pathology-he-auto-augment.
[94] K. Faryna, J. Van Der Laak, and G. Litjens, "Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology." [Online]. Available: https://ssrn.com/abstract=4542792.
[95] N. Marini, et al., Data-driven color augmentation for H&E stained images in computational pathology, J. Pathol. Inform. 14 (2023), https://doi.org/10.1016/j.jpi.2022.100183. Jan.
[96] S. Dammak, M.J. Cecchini, D. Breadner, A.D. Ward, Using deep learning to predict tumor mutational burden from scans of H&E-stained multicenter slides of lung squamous cell carcinoma, J. Med. Imag. 10 (01) (2023), https://doi.org/10.1117/1.jmi.10.1.017502. Feb.
[97] P. Huang et al., "Assessing and enhancing robustness of deep learning models with corruption emulation in digital pathology," Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.20427.
[98] S. Otálora, M. Atzori, V. Andrearczyk, A. Khan, H. Müller, Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology, Front. Bioeng. Biotechnol. 7 (AUG) (2019), https://doi.org/10.3389/fbioe.2019.00198.
[99] N. Bouteldja, D.L. Hölscher, R.D. Bülow, I.S.D. Roberts, R. Coppo, P. Boor, Tackling stain variability using CycleGAN-based stain augmentation, J. Pathol. Inform. 13 (2022), https://doi.org/10.1016/j.jpi.2022.100140. Jan.
[100] Y. Tolkach et al., "Artificial intelligence for tumour tissue detection and histological regression grading in oesophageal adenocarcinomas: a retrospective algorithm development and validation study," 2023. [Online]. Available: www.thelancet.com/.
[101] S. Yang, B.U. Lee, Poisson-Gaussian noise reduction using the hidden Markov model in contourlet domain for fluorescence microscopy images, PLoS ONE 10 (9) (2015), https://doi.org/10.1371/journal.pone.0136964. Sep.
[102] V. Mannam, et al., Real-time image denoising of mixed Poisson–Gaussian noise in fluorescence microscopy images using ImageJ, Optica 9 (4) (2022) 335, https://doi.org/10.1364/optica.448287. Apr.
[103] S. Yang, et al., DeepNoise: signal and noise disentanglement based on classifying fluorescent microscopy images via deep learning, Genomics. Proteomics. Bioinformatics. 20 (5) (2022) 989–1001, https://doi.org/10.1016/j.gpb.2022.12.007. Oct.
[104] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Gene Myers, *Removing structured noise with self-supervised blind-spot networks*.
[105] Y. Zhang et al., "A poisson-gaussian denoising dataset with real fluorescence microscopy images," Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.10366.
[106] E. Demircan-Tureyen, F.P. Akbulut, M.E. Kamasak, Restoring fluorescence microscopy images by transfer learning from tailored data, IEEe Access. 10 (2022) 61016–61033, https://doi.org/10.1109/ACCESS.2022.3181177.
[107] F. Shi, et al., DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images, Phys. Med. Biol. 64 (17) (2019), https://doi.org/10.1088/1361-6560/ab3556. Sep.
[108] N. Gour, P. Khanna, Speckle denoising in optical coherence tomography images using residual deep convolutional neural network, Multimed. Tools. Appl. 79 (21–22) (2020) 15679–15695, https://doi.org/10.1007/s11042-019-07999-y. Jun.
[109] D. Romo-Bucheli, et al., Reducing image variability across OCT devices with unsupervised unpaired learning for improved segmentation of retina, Biomed. Opt. Express. 11 (1) (2020) 346, https://doi.org/10.1364/boe.379978. Jan.
[110] F.G. Venhuizen, et al., Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography, Biomed. Opt. Express. 9 (4) (2018) 1545, https://doi.org/10.1364/boe.9.001545. Apr.
[111] H. Bogunovic, et al., RETOUCH: the Retinal OCT fluid detection and segmentation benchmark and challenge, IEEe Trans. Med. ImAging 38 (8) (2019) 1858–1874, https://doi.org/10.1109/TMI.2019.2901398. Aug.
[112] Z. Ma, D. Feng, J. Wang, H. Ma, Retinal OCTA image segmentation based on global contrastive learning, Sensors 22 (24) (2022), https://doi.org/10.3390/s22249847. Dec.
[113] C. Liu, M. Qiao, F. Jiang, Y. Guo, Z. Jin, Y. Wang, TN-USMA Net: triple normalization-based gastrointestinal stromal tumors classification on multicenter EUS images with ultrasound-specific pretraining and meta attention, Med. Phys. 48 (11) (2021) 7199–7214, https://doi.org/10.1002/mp.15172. Nov.
[114] S. Ren, et al., Preoperative prediction of pathological grading of hepatocellular carcinoma using machine learning-based ultrasomics: a multicenter study, Eur. J. Radiol. 143 (2021), https://doi.org/10.1016/j.ejrad.2021.109891. Oct.
[115] H. Homayoun, et al., Applications of machine-learning algorithms for prediction of benign and malignant breast lesions using ultrasound radiomics signatures: a multi-center study, Biocybern. Biomed. Eng. 42 (3) (2022) 921–933, https://doi.org/10.1016/j.bbe.2022.07.004. Jul.
[116] N. Sirjani, et al., A novel deep learning model for breast lesion classification using ultrasound Images: a multicenter data evaluation, Physica Medica 107 (2023), https://doi.org/10.1016/j.ejmp.2023.102560. Mar.
[117] H. Du, et al., Convolutional networks for the segmentation of intravascular ultrasound images: evaluation on a multicenter dataset, Comput. Methods Programs Biomed. 215 (2022), https://doi.org/10.1016/j.cmpb.2021.106599. Mar.
[118] N.C.F. Codella, et al., Deep learning ensembles for melanoma recognition in dermoscopy images, IBM. J. Res. Dev. 61 (4) (2017), https://doi.org/10.1147/JRD.2017.2708299. Jul.
[119] R. Azad, M.T. Al-Antary, M. Heidari, D. Merhof, TransNorm: transformer provides a strong spatial normalization mechanism for a deep segmentation model, IEEe Access. 10 (2022) 108205–108215, https://doi.org/10.1109/ACCESS.2022.3211501.
[120] Z. Yu, et al., Melanoma recognition in dermoscopy images via aggregated deep convolutional features, IEEE Trans. Biomed. Eng. 66 (4) (2019) 1006–1016, https://doi.org/10.1109/TBME.2018.2866166. Apr.
[121] A. Gong, X. Yao, W. Lin, Dermoscopy image classification based on StyleGANs and decision fusion, IEEe Access. 8 (2020) 70640–70650, https://doi.org/10.1109/ACCESS.2020.2986916.
[122] M.S. Ali, M.S. Miah, J. Haque, M.M. Rahman, M.K. Islam, An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models, Mach. Learn. Appl. 5 (2021) 100036, https://doi.org/10.1016/j.mlwa.2021.100036. Sep.
[123] C. Xin, et al., An improved transformer network for skin cancer classification, Comput. Biol. Med. 149 (2022), https://doi.org/10.1016/j.compbiomed.2022.105939. Oct.
[124] H.K. Gajera, D.R. Nayak, M.A. Zaveri, A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features, Biomed. Signal. Process. Control 79 (2023), https://doi.org/10.1016/j.bspc.2022.104186. Jan.
[125] K. Zafar, et al., Skin lesion segmentation from dermoscopic images using convolutional neural network, Sensors (Switzerland) 20 (6) (2020), https://doi.org/10.3390/s20061601. Mar.
[126] K. Behara, E. Bhero, J.T. Agee, Skin lesion synthesis and classification using an improved DCGAN classifier, Diagnostics 13 (16) (2023), https://doi.org/10.3390/diagnostics13162635. Aug.
[127] C. Barata, J.S. Marques, M.E. Celebi, Improving dermoscopy image analysis using color constancy, in: 2014 IEEE International Conference on Image Processing, ICIP 2014 19, 2014, pp. 3527–3531, https://doi.org/10.1109/ICIP.2014.7025716.
[128] C. Barata, M.E. Celebi, J.S. Marques, Improving dermoscopy image classification using color constancy, IEEe J. Biomed. Health Inform. 19 (3) (2015) 1146–1152, https://doi.org/10.1109/JBHI.2014.2336473. May.
[129] Q. Abbas, I.F. Garcia, M.Emre Celebi, W. Ahmad, Q. Mushtaq, A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images, Skin Res. Technol. 19 (1) (2013), https://doi.org/10.1111/j.1600-0846.2012.00670.x. Feb.
[130] B. Hewitt, M.H. Yap, J. Ng, M. Goyal, The effect of color constancy algorithms on semantic segmentation of skin lesions, in: SPIE-Intl Soc Optical Eng, 2019, p. 25, https://doi.org/10.1117/12.2512702. Mar.
[131] Olga Cherepkova, Jon Yngve Hardeberg, Enhancing Dermoscopy Images to Improve Melanoma Detection, IEEE, 2018.

[132] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification, IEEe Trans. Med. ImAging 38 (9) (2019) 2092–2103, https://doi.org/10.1109/TMI.2019.2893944. Sep.
[133] J. van de Weijer, T. Gevers, A. Gijsenij, Edge-based color constancy, IEEE Trans. Image Process. 16 (9) (2007) 2207–2214, https://doi.org/10.1109/TIP.2007.901808. Sep.
[134] Y. Yuan, Y.C. Lo, Improving Dermoscopic image segmentation with enhanced convolutional-deconvolutional networks, IEEe J. Biomed. Health Inform. 23 (2) (2019) 519–526, https://doi.org/10.1109/JBHI.2017.2787487. Mar.
[135] M. Goyal, A. Oakley, P. Bansal, D. Dancey, M.H. Yap, Skin lesion segmentation in Dermoscopic images with ensemble deep learning methods, IEEe Access. 8 (2020) 4171–4181, https://doi.org/10.1109/ACCESS.2019.2960504.
[136] G.D. Finlayson and E. Trezzi, “Shades of gray and Colour constancy”.
[137] M. Salvi, et al., DermoCC-GAN: a new approach for standardizing dermatological images using generative adversarial networks, Comput. Methods Programs Biomed. 225 (2022), https://doi.org/10.1016/j.cmpb.2022.107040. Oct.
[138] A. Galdran et al., “Data-driven color augmentation techniques for deep skin image analysis,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.03702.
[139] F. Veronese, et al., The role in Teledermoscopy of an inexpensive and Easy-to-Use smartphone device for the classification of three types of skin lesions using convolutional neural networks, Diagnostics 11 (3) (2021) 451, https://doi.org/10.3390/DIAGNOSTICS11030451. *2021, Vol. 11, Page 451*Mar.
[140] F. Branciforti, et al., Impact of artificial intelligence-based color constancy on dermoscopical assessment of skin lesions: a comparative study, Skin Res. Technol. 29 (11) (2023), https://doi.org/10.1111/srt.13508. Nov.
[141] M. Salvi, F. Branciforti, F. Molinari, K.M. Meiburger, Generative models for color normalization in digital pathology and dermatology: advancing the learning paradigm, Expert. Syst. Appl. 245 (2024) 123105, https://doi.org/10.1016/j.eswa.2023.123105. Jul.
[142] A. Papadopoulos, D.I. Fotiadis, L. Costaridou, Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques, Comput. Biol. Med. 38 (10) (2008) 1045–1055, https://doi.org/10.1016/j.compbiomed.2008.07.006. Oct.
[143] S.K. Kang, D. Kim, S.A. Shin, Y.K. Kim, H. Choi, J.S. Lee, Fast and accurate amyloid brain PET quantification without MRI using deep neural networks, J. Nucl. Med. 64 (4) (2023) 659–666, https://doi.org/10.2967/jnumed.122.264414. Apr.
[144] R.D. Brook, et al., Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the american heart association, Circulation 121 (21) (2010) 2331–2378, https://doi.org/10.1161/CIR.0b013e3181dbece1. Jun. 01.
[145] K. Sirinukunwattana, et al., Gland segmentation in colon histology images: the glas challenge contest, Med. Image Anal. 35 (2017) 489–502, https://doi.org/10.1016/j.media.2016.08.008. Jan.
[146] M. Veta, et al., Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge, Med. Image Anal. 54 (2019) 111–121, https://doi.org/10.1016/J.MEDIA.2019.02.012. May.
[147] G. Litjens, et al., 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset, Gigascience 7 (6) (2018), https://doi.org/10.1093/gigascience/giy065. Oxford University PressJun. 01.
[148] P. Bándi, et al., From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge, IEEe Trans. Med. ImAging 38 (2) (2019) 550–560, https://doi.org/10.1109/TMI.2018.2867350. Feb.
[149] Á.E. Esteban, M. López-Pérez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes, Comput. Methods Programs Biomed. 178 (2019) 303–317, https://doi.org/10.1016/j.cmpb.2019.07.003. Sep.
[150] K. Tomczak, P. Czerwińska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, Wspolczesna Onkologia 1A (2015) A68–A77, https://doi.org/10.5114/wo.2014.47136. Termedia Publishing House Ltd.
[151] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2016) 1455–1462, https://doi.org/10.1109/TBME.2015.2496264. Jul.
[152] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEe Trans. Med. ImAging 36 (7) (2017) 1550–1560, https://doi.org/10.1109/TMI.2017.2677499. Jul.
[153] G. Aresta, et al., BACH: grand challenge on breast cancer histology images, Med. Image Anal. 56 (2019) 122–139, https://doi.org/10.1016/j.media.2019.05.010. Aug.
[154] Y.J. Kim, et al., PAIP 2019: liver cancer segmentation challenge, Med. Image Anal. 67 (2021), https://doi.org/10.1016/j.media.2020.101854. Jan.
[155] M. Li et al., “OCTA-500: a retinal dataset for optical coherence tomography angiography study”, Accessed: Jan. 27, 2024. [Online]. Available: https://ieee-dataport.org/open-access/octa-500.
[156] N. Banić, S. Lončarić, Smart light random memory sprays Retinex: a fast Retinex implementation for high-quality brightness adjustment and color correction, J. Opt. Soc. Am. A 32 (11) (2015) 2136, https://doi.org/10.1364/josaa.32.002136. Nov.
[157] Y. Peng, N. Wang, Y. Wang, M. Wang, Segmentation of dermoscopy image using adversarial networks, Multimed. Tools. Appl. 78 (8) (2019) 10965–10981, https://doi.org/10.1007/s11042-018-6523-2. Apr.
[158] R.J. Chen, et al., Algorithmic fairness in artificial intelligence for medicine and healthcare, Nat. Biomed. Eng. 7 (6) (2023) 719–742, https://doi.org/10.1038/s41551-023-01056-8. *2023 7:6*Jun.
[159] M. Salvi, et al., Multi-modality approaches for medical support systems: a systematic review of the last decade, Inf. Fusion 103 (2024) 102134, https://doi.org/10.1016/j.inffus.2023.102134. Mar.
[160] R.S. Antunes, C.A. Da Costa, A. Küderle, I.A. Yari, B. Eskofier, Federated learning for healthcare: systematic review and architecture proposal, ACM. Trans. Intell. Syst. Technol. 13 (4) (2022), https://doi.org/10.1145/3501813. May.
[161] A.B.E. Attia, et al., A review of clinical photoacoustic imaging: current and future trends, Photoacoustics 16 (2019), https://doi.org/10.1016/j.pacs.2019.100144. Elsevier GmbHDec. 01.
[162] F. Pérez-Bueno, et al., Blind color deconvolution, normalization, and classification of histological images using general super Gaussian priors and Bayesian inference, Comput. Methods Programs Biomed. 211 (2021), https://doi.org/10.1016/j.cmpb.2021.106453. Nov.