Artificial Intelligence in the Life Sciences 4 (2023) 100083

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

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

Deep neural network architectures for cardiac image segmentation

Jasmine El-Taraboulsi [a], Claudia P. Cabrera [b,c], Caroline Roney [d], Nay Aung [b,c,e,*]

[a] Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
[b] Barts and The London School of Medicine and Dentistry, William Harvey Research Institute, Queen Mary University of London, London, United Kingdom
[c] National Institute for Health and Care Research Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, United Kingdom
[d] School of Engineering and Materials Science, Queen Mary University of London, London, United Kingdom
[e] Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, United Kingdom

ARTICLE INFO

Keywords:
Deep learning
Cardiac imaging
Image segmentation
Machine Learning

ABSTRACT

Imaging plays a fundamental role in the effective diagnosis, staging, management, and monitoring of various cardiac pathologies. Successful radiological analysis relies on accurate image segmentation, a technically arduous process, prone to human-error. To overcome the laborious and time-consuming nature of cardiac image analysis, deep learning approaches have been developed, enabling the accurate, time-efficient, and highly personalised diagnosis, staging and management of cardiac pathologies.

Here, we present a review of over 60 papers, proposing deep learning models for cardiac image segmentation. We summarise the theoretical basis of Convolutional Neural Networks, Fully Convolutional Neural Networks, U-Net, V-Net, No-New-U-Net (nnU-Net), Transformer Networks, DeepLab, Generative Adversarial Networks, Auto Encoders and Recurrent Neural Networks. In addition, we identify pertinent performance-enhancing measures including adaptive convolutional kernels, atrous convolutions, attention gates, and deep supervision modules.

Top-performing models in ventricular, myocardial, atrial and aortic segmentation are explored, highlighting U-Net and nnU-Net-based model architectures achieving state-of-the art segmentation accuracies. Additionally, key gaps in the current research and technology are identified, and areas of future research are suggested, aiming to guide the innovation and clinical adoption of automated cardiac segmentation methods.

1. Introduction

1.1. Cardiac image segmentation

Cardiovascular diseases are the prevailing cause of mortality and morbidity globally, encompassing diverse pathologies affecting the heart and its arteries [1]. Profuse advancements in medical imaging over the last decade form the cornerstone of the clinical diagnosis, staging and management of cardiac diseases [2]. As the clinical environment shifts towards personalised medicine, individualised quantification and analysis of cardiac structure and function is paramount [2]. Nonetheless, the unique and complex geometry of cardiac substructures, coupled with the heart's location and orientation, continuous motion, and vast anatomical variability makes imaging a persistent challenge [2].

Cardiac imaging techniques can be broadly categorised into invasive and non-invasive modalities. Mainstay non-invasive imaging modalities include echocardiography (Echo), nuclear scintigraphy myocardial perfusion imaging, Cardiac Magnetic Resonance (CMR), and Computed Tomography (CT) [3]. In contrast, invasive imaging techniques predominantly include Coronary Angiography (CA) using cardiac catheterisation, intravascular ultrasound and intravascular optical coherence tomography, occasionally in combination with non-invasive modalities [3]. Each modality has its own advantages and disadvantages and can be used individually or in combination with one another, contingent on clinical indications and context [3]. The key functions of each modality are summarised in Table 1.

Once imaging has been obtained with the modality of choice, the output requires analysis and reporting by a cardiologist with imaging expertise, or a radiologist. This analysis extracts quantitative and qualitative features from cardiac images, enabling precise conclusions that are not apparent to the human eye to be drawn (Fig. 1). Examples of estimated metrics include myocardial scarring quantification, discerning the distribution of myocardial thickening, computing Left Ventricular (LV) and Right Ventricular (RV) volume, and computation of

* Corresponding author at: William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, United Kingdom.
E-mail address: n.aung@qmul.ac.uk (N. Aung).

https://doi.org/10.1016/j.ailsci.2023.100083
Received 30 June 2023; Received in revised form 6 August 2023; Accepted 6 August 2023
Available online 9 August 2023

Table 1
Clinical applications of Echo, Nuclear Imaging, CT, CMR, and coronary angiography [3–5].

Imaging Modality	Clinical Significance	Advantages	Disadvantages
Echo	• Images chamber size, ventricular wall thickness, structural abnormalities valvular dysfunction (including vegetations and thrombi), LV contractility and EF • 3D Echo assesses chamber volume, structural abnormalities, valvular disorders • Strain rate imaging determines regional/global deformity of heart muscles • Stress echo detects myocardial ischaemia • Doppler flow determines velocity of blood flow and visualises blood flow	• Highly accessible • Low costs • Can be used for pre- and post-procedure follow-up • No radiation exposure • Non-invasive	• High degree of operator dependence • High degree of interobserver variability • Patient-factor related image quality issues (e.g., inadequate acoustic window)
Nuclear Imaging	• Stress testing to assess for ischaemia • Suspected CAD, demonstrating areas of hypo perfusion using radioactive tracers • Can assess myocardial viability and degree of scarring	• Low degree of inter-operator dependence • High reproducibility of LVEF • Non-invasive	• Radiation exposure • Low availability • Relative high costs
CT	• Contrast is often used to differentiate between cardiac chambers and vascular structures • Calcium scoring for detecting coronary artery calcification • Ability to evaluate aortic pathology • Perfusion and fractional flow reserve • Produce 3D images of the heart	• High signal: noise ratio • Superior spatial resolution • High accuracy • High reproducibility • Non-invasive	• Radiation exposure • High costs • Risk of contrast-induced nephropathy (if contrast is used) • Lower availability
CMR	• Quantification of cardiac volume and function • First-pass perfusion imaging to assess myocardial ischaemia • Tissue characterisation with Late Gadolinium-Enhancement (LGE) CMR to detect replacement fibrosis, and mapping techniques (T1, T2, T2*, ECV) to assess diffuse fibrosis, oedema, iron overload and infiltrative cardiomyopathies • Phase contrast imaging and 4D flow for assessment of flow, valves and myocardial performance • Strain imaging (tagging, feature-tracking) • Large vasculature visualisation	• High accuracy • High reproducibility • Low degree of inter-operator dependence • Non-invasive	• Lower availability • Presence of metallic prostheses, cardiovascular implantable electronic devices (CIEDs), although more recently patients with CIEDs have been safely scanned with appropriate precautions • Risk of gadolinium contrast (nephrogenic systemic fibrosis) • Lower temporal resolution compared to echo
Coronary Angiography	• Visualisation of coronary anatomy in real-time • Measure of hemodynamic pressures	• Can be therapeutic by treating coronary lesions	• Peri-procedural risks (MI, stroke, arrhythmias) • Risk of contrast-induced nephropathy • Invasive

Ejection Fraction (EF), amongst several others [6]. Thus, imaging analysis facilitates the diagnosis, staging, monitoring, and management of a vast array of disorders, by providing objective and quantitative evidence for pathology [6].

Cardiac imaging analysis involves image acquisition, image segmentation and Region-of-Interest (ROI) definition [6]. Image segmentation, the focus of this paper, involves classifying the pixels within an image of an organ or substructure based on their specific features [7]. Following this, feature extraction, selection and classification may occur to yield the desired outcome [6]. Thus, accurate and efficient image segmentation is critical to successful cardiac image interpretation, enabling advanced structural processing, 3D reconstruction, and cardiac motion analysis [7]. Cardiac image segmentation constitutes a technically challenging task, significantly influenced by image quality and artefact [6]. The segmentation approach falls into three broad categories: manual, semi-automatic and automatic segmentation [6].

Manual segmentation, the current gold-standard, requires expert cardiac radiologists or imaging cardiologists to analyse each slice of two-dimensional (2D), or three-dimensional (3D) images and annotate the ROI (Fig. 2) [8]. This classically involves contouring to outline endocardial and epicardial borders, and subsequently calculating cardiac function and myocardial mass [6]. The evident advantage of manual segmentation is that it utilises expert knowledge, often comprising the ground-truth in segmentation tasks [8]. However, a clear drawback is the highly time-consuming and laborious nature of the task, typically requiring up to a third of analysis time per cardiac study, in addition to extensive inter-observer variability reducing segmentation precision [8, 9]. For example, pericardial fat and trabeculations may cause the RV's borders to appear fuzzy on CMR images, coupled with image artefact secondary to poor breath-hold, irregular heart rhythms or potential prosthetic materials, making RV segmentation particularly challenging and prone to inaccuracies [10].

Semi-automatic segmentation relies on an automated segmentation framework, followed by manual review by an in-field expert to adjust or correct the segments [4]. Prior to Deep Learning (DL) algorithms, this was frequently used to undertake volumetric quantification of the cardiac structures [6]. Examples of semi-automated segmentation approaches include image-driven algorithms, probabilistic atlases, fuzzy clustering and anatomical-based landmarks [4]. The major disadvantage of semi-automated segmentation is its reliance on manual initialisation of segmentation, thus, still prone to inter-observed variability [4].

In contrast, fully automatic segmentation requires no manual input, and can be applied to a diverse range of scenarios including whole heart, myocardial scarring, and coronary artery segmentation [6]. Automating image segmentation is more time-efficient than traditional methods, permitting hundreds to thousands of images to be processed in seconds to minutes [11]. While both non-DL and DL-based techniques can be applied to automatic image segmentation, DL-based approaches are highly accurate and robust to anatomical variations, despite relying on less computational power and minimal prior knowledge at the time of inference [11]. Significant advancements have been made in the field of automated cardiac image segmentation, with DL models at times outperforming humans (e.g., in estimation of myocardial thickness).

1.2. *Machine learning*

Machine Learning (ML) approaches have transformed data analysis and information processing, creating programs that learn from experience without requiring a hard-coded "knowledge base", using advanced pattern recognition to solve experience-based, real-world problems [9].

Traditional ML methods such as logistic regressions and naïve Bayes Algorithms succeed in simple classification tasks, contingent on pre-defined data representations as inputs, but demonstrate declining performance when features are unknown [9]. Representation learning constitutes an ML method used to automate representation mapping and feature detection, utilising encoder-decoder (AutoEncoder) functions to convert inputs into new representations [11]. However, emulating the decision-making process of the human brain requires integrating visible and non-visible features, while simultaneously ranking their importance, automatically overlooking irrelevant variables [9]. Defining such a representation is exceedingly complex, thus evoking the establishment of DL, a type of ML characterised by hierarchical nested layers, with multiple interconnected representations aiming to map abstract and complex concepts [9,12].

There are two general categories of ML models: supervised and unsupervised. The former encompasses a dataset where each example has a labelled target, teaching the machine what to do to achieve the correct response. The latter represents an unlabelled dataset where the algorithm must "teach itself" the full probability distribution of the dataset [9]. Semi-supervised ML models comprise a middle ground, where only some examples contain a target within a dataset.

The classical perceptron, proposed by Rosenblatt in 1958, initialises the concept of numerical weights, scalar values attached to each feature reflecting their importance [13]. This has since become quintessential to DL architectures, creating the foundation of the multilayer perceptron (MLP), the most basic form of Neural Networks (NNs) [13,14]. MLPs consist of an input and output layer connected by a variety of hidden layers, with each layer composed of multiple neurons (Fig. 3). The input layer is a vector of predictable values, with the number of neurons equivalent to the number of predictable values [15]. This layer standardises the range of the vector values and distributes them across the neurons within the hidden layers [15]. Additionally, a bias is projected onto each hidden layer, this is a constant that is added to the product of the input and weight, aiming to balance the results [15]. The hidden layer multiplies the projected inputs by randomly initialised weights, the resultant sum is passed through an activation function contributing to the calculation of an output [15,16]. The function of the hidden layers is influenced by the learning algorithm, where the model must determine how the layers can produce the most accurate representation of the objective function [9]. Hence, network behaviour is not pre-specified by the training data, unlike in traditional ML methods [9]. The dimensionality of the hidden layer influences model "width," while the number of hidden layers determines model "depth" [9]. When defining the number of hidden layers within a model, one layer is sufficient unless the available data has discontinuities [17]. However, determining the number of neurons within a hidden layer requires striking a meticulous balance, where too few neurons make it impossible to model complex relationships, while too many neurons put the model at risk of overfitting [15]. Overfitting is a phenomenon that occurs when the model is too complex for its data, giving disproportionate importance to noisy, insignificant data, generating very accurate results on training data, but performing poorly on unseen test sets (low generalisation) [15]. The values obtained from each hidden layer are multiplied by weights, and summed, producing a vector that passes through a transformation function, generating output values [15]. Subsequently, output values are either utilised for back-propagation during model training, or for decision-making during testing [16]. When training an MLP the primary objective is to establish a set of weights that when multiplied produce output values closest to the target value [15].

Typically, the available data is split into a training set and a testing set, where the model's performance on the unseen testing set can be utilised for evaluation of generalisation error and accuracy [18]. The training set is split further into two subsets, one is used to learn parameters (typically 80% of training set), and the other is used as the validation set (typically 20% of training set), measuring generalisation error and permitting the updating of hyperparameters during training [18].

In the context of image segmentation, the image would be the input, with each pixel representing a feature (Fig. 4a). The principal outcome is to build a NN that can accurately predict the class of each pixel using examples from both the training and test set [9]. Hence, a balance between under-fitting and over-fitting must be struck, equating the model's capacity with the task's complexity [9]. Various features of NNs are adapted to optimise training outcomes. Firstly, loss functions (also known as model cost or error) are fundamental to ML, measuring the discrepancy between the model's estimation of an input to an output, and the ground-truth value (distance between predicted value and actual value) [18]. Minimising the loss function is achieved using optimisation algorithms, teaching models how to adjust their parameters (model weight and capacity), aiming to reach a point of convergence [18]. Back-propagation is the process by which weights are individually updated to reduce the loss function following every iteration; the gradient of the loss function of each weight is computed and adjusted accordingly [19]. Hyperparameters encompass manually adjustable

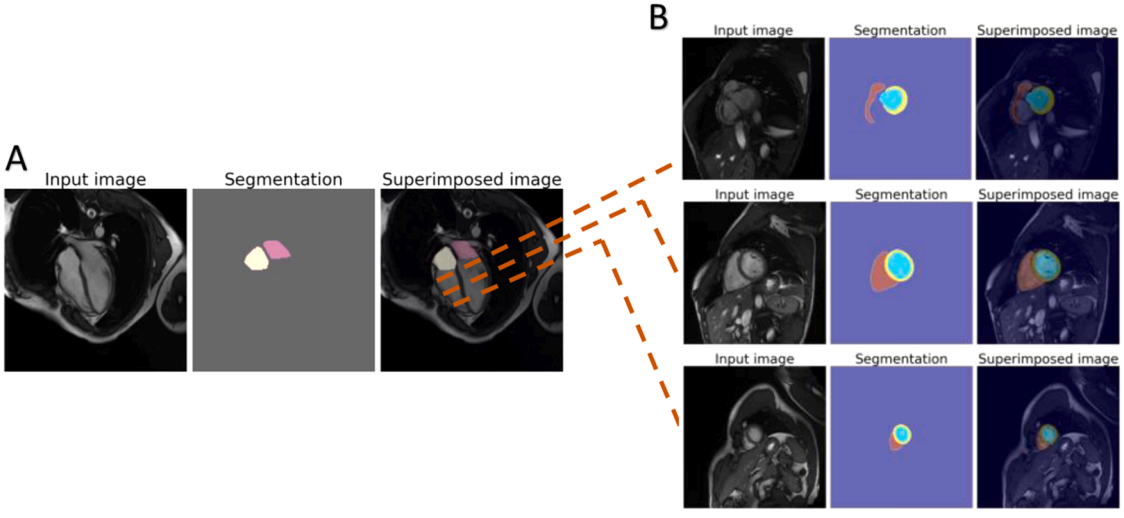

Fig. 1. Exemplary automated atrial and ventricular segmentation in long- and short-axis CMR images obtained from M&Ms 2 dataset (https://www.ub.edu/mnms-2/) using a nnU-Net-based architecture. (A) Demonstrates atrial segmentation in the long-axis 4 chamber view; (B) Demonstrates ventricular segmentation in the short-axis view for the same subject as panel A.

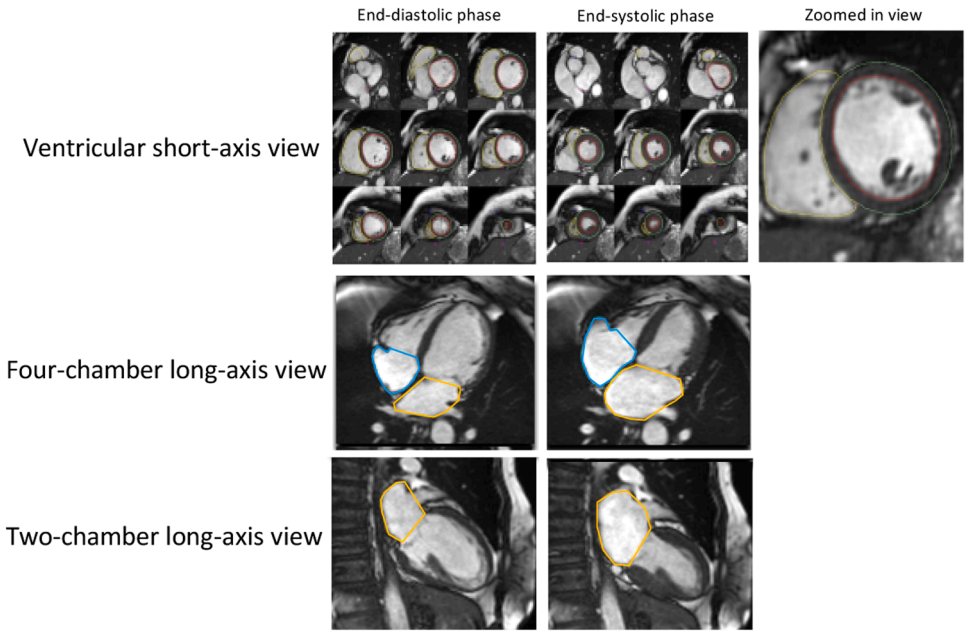

Fig. 2. Manual segmentation process demonstrating delineation of the cardiac chambers by region-of-interest tracing. End-diastolic, end-systolic and zoomed-in views are presented across various CMR views.

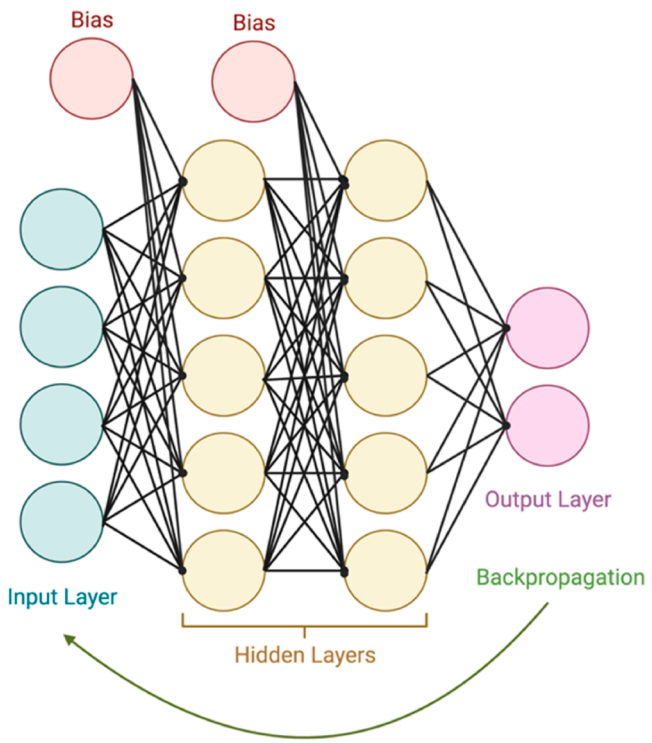

Fig. 3. Multilayer Perceptron network comprised of an input layer, hidden layers with corresponding biases projected, an output layer, and back-propagation for model training.

settings, such as the degree of a polynomial's regression [18]. As such, hyperparameters may control model capacity, and must therefore be adjusted to prevent over-fitting [18].

This review aims to provide a comprehensive investigation of the status of DL architectures in cardiac image segmentation by analysing over 60 articles published between January 1, 2019, and January 13, 2023. First, we present a theoretical overview of the predominant NN architectures used in cardiac image segmentation, advanced building blocks that can be applied to enhance results, and commonly employed loss functions. Then, we describe the methodology of the literature review and present a simplified version of the results. Finally, we summarise the top performing NN architectures across various cardiac segmentation tasks, delineate key challenges currently encountered by state-of-the-art segmentation models, and suggest areas of future investigation.

1.3. Neural network architectures

In the next section, we present the theoretical frameworks of commonly used NN architectures in image segmentation tasks: Convolutional Neural Networks (CNN), Fully Convolutional Networks (FCN), U-Net, Versatile Network (V-Net), No-New-Net (nnU-Net), Transformer, DeepLab, Generative Adversarial Networks (GAN), AutoEncoders (AE) and Recurrent Neural Networks (RNN).

1.3.1. Convolutional neural networks (CNN)

CNN are the most prevalent NN architecture in image classification, detection, and segmentation tasks [20]. These networks are characterised by a 'grid-like' topology, consisting of an input and output layer, with functional layers in between [20]. The functional layers typically include convolutional layers, pooling layers, and fully connected layers [20].

Convolutional layers, the defining characteristic of a CNN, consists of an input, a second argument (also referred to as the kernel), and an output (feature map) [9]. The input and kernel are multidimensional arrays of data and parameters respectively [9]. The convolutional kernel is classically followed by a normalisation layer, and a non-linear activation function, such as a rectified linear activation function (ReLu), releasing an output corresponding to the input (Fig. 4c) [9]. All outputs are decimated through pooling layers, usually down-sampling by a factor of two, aiming to optimise efficiency, accuracy, and generalisability by excluding redundant features [20]. Max-pooling operations determine the highest value within a section of a feature map and use this to create a down-sampled feature map (Fig. 4d). Consequently, the output becomes invariant to minimal changes in the input. This is key in certain image segmentation tasks where the presence of a feature is

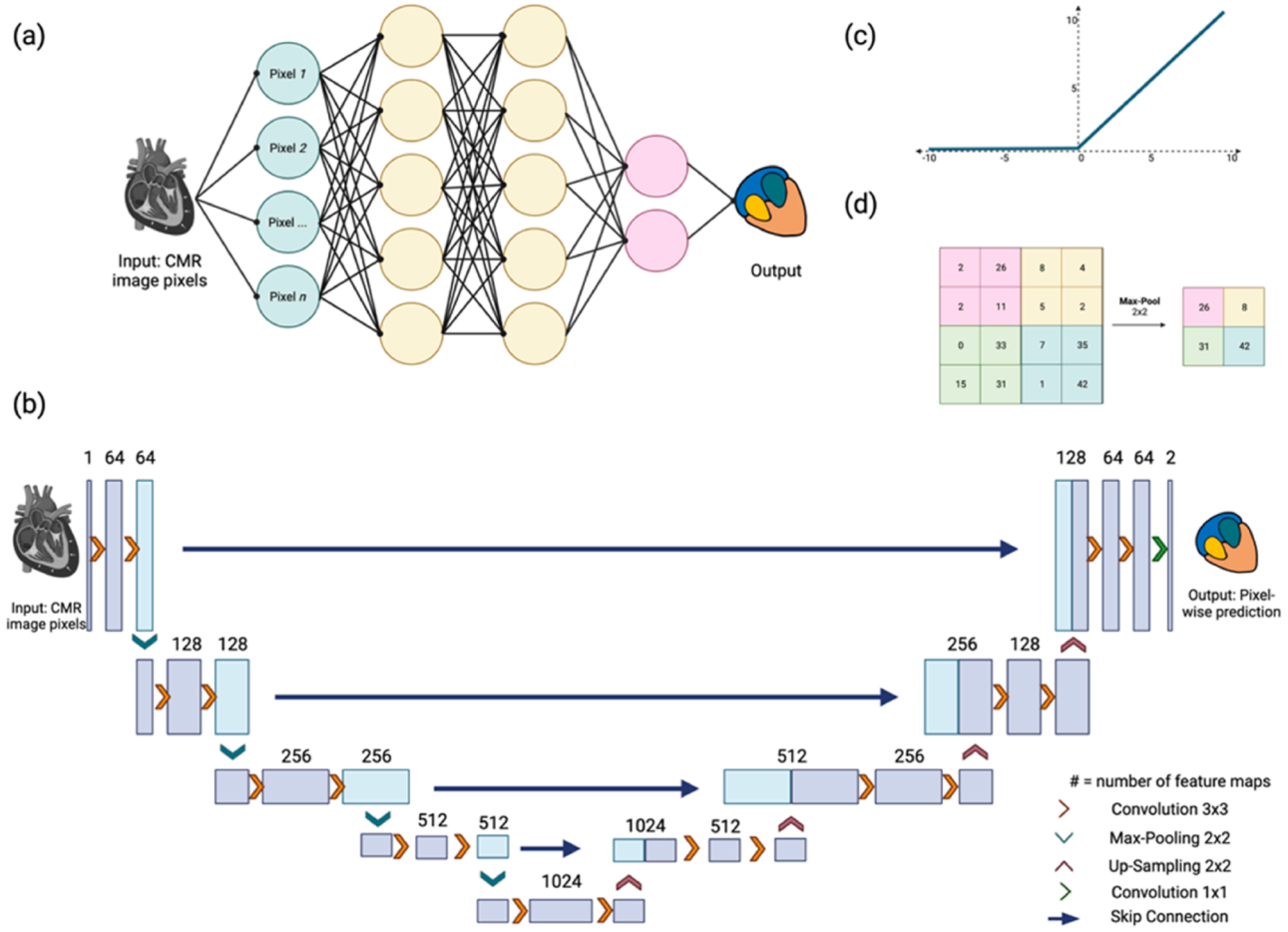

Fig. 4. (a) Basic feedforward neural network using a CMR image as input; (b) U-Net model designed for CMR segmentation, comprised of contracting and expansive pathways using max-pooling and up-sampling to generate pixel-wise predictions; (c) Graphical representation of the ReLu activation function graph; (d) 2 × 2 Max-Pooling operation.

more important than its location, significantly increasing computational efficiency [9]. For example, when segmenting cardiac images, at times it may be important to merely recognise the location of the major heart chambers, however, it might not be necessary to view these structures with optimal pixels [9]. Furthermore, pooling is indispensable to image processing tasks that deal with inputs of disparate sizing [9]. Next, the fully connected layer establishes the features most vital to successful prediction, and thus decreases the dimensionality of features from the preceding layer [20]. Finally, a fix-sized vector is produced as the model output [20].

There are three key benefits that arise from convolutional layers, namely sparse interactions, parameter sharing and equivariant representations [9]. Firstly, sparse interactions are a feature of CNNs that reduce the kernel to a size smaller than the input, differing from classical NN architectures as each output unit does not need to interact with each input [9]. This poses an apparent advantage in image segmentation, as instead of storing the millions of pixels that may be associated with an input image, only meaningful features are processed in kernels composed of tens or hundreds of pixels [9]. As a result, processing efficiency is significantly increased, while running time is reduced [9]. Additionally, in deep CNNs, reducing kernel size allows the receptive field to increase by augmenting the number of convolutional layers, this causes an indirect interaction with a greater proportion of inputs enabling complex, multifactorial correlations to be captured [9,20]. Secondly, parameter sharing (or tied weights) causes the value of one input's weight factor to be applied to all other weight values [9]. In other words, all constituents of a kernel are used throughout every input position, allowing only one set of parameters to be learnt for all locations [9]. Equivariance is a product of parameter sharing, where a change to the input produces the same transformation to the output [9]. This is advantageous to image processing tasks as the first convolutional layer typically detects image edges [9]. Since, images typically share borders, equivariance enables effective parameter sharing [9].

CNN architectures specific to image segmentation utilise a patch-based approach, where an image is divided into several smaller patches, and the model is trained to predict the class-label of each patch's central pixel [20]. This generates an inherent inefficiency as there is significant overlap in the image area covered by each patch, however, despite this, the model is required to train on each patch individually [20]. Furthermore, localisation accuracy is spared at the cost of maintaining context. Larger patches oblige a greater number of pooling layers, compromising the localisation accuracy but preserving context, while smaller patches have a greater localisation accuracy but lack context, as they cover smaller image areas [21]. Thus, traditional CNNs are typically utilised for object localisation in cardiac image segmentation. More recently, modifications to traditional CNNs have been proposed that enable complete pixel-wise segmentation [20].

1.3.2. Fully convolutional networks (FCN)

FCN encompass a ground-breaking variant of CNNs, designed to undertake pixel-to-pixel prediction tasks without systematic inefficiency [22]. Long et al. developed the FCN for semantic segmentation, aiming to overcome shortcomings associated with traditional CNNs [88]. FCNs do not contain any "dense" or fully connected layers, thus, exclusively

consisting of convolutional layers [22,23]. Every data point with an FCN consists of three dimensions, *h x w x d,* where h and w represent height and width (spatial dimensions), while d represents the feature/channel dimension [22]. FCN input images can have variable sizes that will be encoded into feature representations, and then decoded using spatial information via a sequence of up-sampling (deconvolution) and convolutional layers [20]. Up-sampling increases the magnitude of minority classes by adding duplicate data points to that class, enabling a balanced data set [89]. In contrast to the traditional CNN patch-based approach, FCNs can train and make predictions on entire images [20]. Nonetheless, FCNs encoder-decoder approach causes the elimination of notable features and contextual information during the pooling layers [20]. Hence, updates to traditional FCNs have been proposed, aiming to transmit features between encoding and decoding layers, preserving spatial context, and improving segmentation accuracy [20]. most widely employed architecture for biomedical image segmentation is the U-Net, a variant of the FCN [20]. This model's U-shaped architecture incorporates 23 convolutional layers split into a contracting (encoder) pathway and an expansive (decoder) pathway (Fig. 4b) [21]. The contracting path is like traditional CNNs, composed of two convolution layers with associated ReLu and max pooling operations, where the number of features is doubled at each down-sampling step [21]. The expansive path uses up-sampling at each step followed by a convolution, halving the number of features [21]. All feature maps are then cropped due to the loss of border pixels at each convolution [21]. Concatenating skip connections are present between contracting and expansive pathways, integrating each encoder's feature map within the corresponding decoder. These enable recovery of spatial context, increasing segmentation precision [20]. The cropped feature maps from the contracting pathway are thus connected to the expansive pathway, and projected into two convolutions, each proceeded by a ReLu function [21]. The final layer is marked by a single convolution that maps the output [21].

Hence, U-Net supplements CNN's network using up-sampling layers to maintain feature quality and output precision [21]. Moreover, during up-sampling, U-Net has numerous feature channels enabling the transmission of high-resolution information across layers [21]. Additionally, this architecture uses an "overlap-tile strategy" to tackle large images while minimally impacting processing power [21].

2D and 3D U-Net models are powerful variants of traditional U-Nets, utilising similar architectures with alterations to the kernel size and convolutional layers to reflect the dimensionality of the input [24]. The fundamental difference between both data types is that 2D models train and make predictions based on a single slice, whereas 3D models can make inter-slice predictions [24]. While this enables more complex and insightful segmentation, 3D models have an increased cost of computing, requiring patch-based processing [24]. In general, both 2D and 3D U-Nets are effective at biomedical image segmentation, with their accuracy varying with task complexity [24].

1.3.3. Versatile network (V-Net)

V-Net is a deep learning methodology used for semantic segmentation, designed to overcome the deep and wide nature of CNN layers [25]. This architecture employs a reversible mechanism and asymmetrical convolutions maintaining image size and quality [25]. As a result, V-Net can train high-quality images on a single GPU [25]. This model compiles Contextual Pyramid Pooling modules and versatile modules [25].

1.3.4. No-New-Net (nnU-Net)

Isensee's nnU-Net provides a powerful segmentation solution that can automatically perform data pre-processing, hyperparameter and parameter optimisation, and output post-processing using a classical U-Net encoder-decoder [26]. This method groups knowledge into fixed, rule-based, or empirical parameters, building "dataset fingerprints" that capture key features of a dataset representation which are dependant on "pipeline fingerprints," comprising the spectrum of choices available during methodology design [26]. Interdependent heuristic rules are built secondary to the dependencies, available for prompt deployment and application without increasing computational load [26]. Three distinctive U-Net models are then automatically generated, a 2D model, 3D model, and a 3D cascaded U-Net, and the best-performing is selected [26]. Hence, nnU-Net proposes an end-to-end automated segmentation methodology with state-of-the-art performance standards.

1.3.5. Transformer

Transformer encompasses a DL model that was initially developed for natural language processing but has recently been introduced to the image processing domain [27]. Transformers alone do not employ CNN-based architectures; however, modifications have been performed leading to the creation of TransUNet [27]. This model utilises a Vision Transformer (VIT) as the encoder and a CNN as the decoder [27]. The VIT deploys the transformer architecture onto fix-sized patches present within the image, the linear embeddings provided by these patches are then input into a Transformer model [28]. Thus, TransUNet overcomes the lack of spatial context produced by only VIT models [27].

Another novel variant of the transformer is the Shifted Windows (Swin) Transformer, a hierarchical Transformer characterised by non-overlapping windows that still permit cross-window connections [29]. Swin Transformers divide input images into non-overlapping patches, where each patch is a "token", Swin Transformer blocks and linear embeddings are applied onto each token in stage 1 of the model's architecture [29]. Stage 2, 3 and 4 are characterized by patch merging layers, concatenating the features of neighbouring patches, followed by Swin Transformer blocks [29]. Each stage is characterized by a unique output resolution.

The Swin Transformer block replaces the traditional Transformer's multi-head self-attention module with a shifted window self-attention module [29]. Resultantly, Swin Transformer limits computation to the non-overlapping windows, improving efficiency while enabling processing at various scales and image sizes [29]. This model offers a general backbone for image classification tasks, differentiating from other vision Transformers' low resolution [29].

1.3.6. DeepLab

DeepLab delineates an alternative semantic segmentation model utilising an encoder-decoder based architecture [30]. DeepLab's encoder is composed of a CNN model, and its decoder utilises up-sampling to reconstruct the output [30]. This approach aberrates from deep CNNs as the numerous max-pooling layers within CNN's model architecture ultimately decreases feature map spatial resolution [17]. Although models such as U-Net have integrated de-convolutional layers (up-sampling layers) to maintain spatial resolution, DeepLab utilises an alternative mechanism entitled atrous convolution [17]. Atrous convolution is analogous to down-sampling layers in CNN models, however, it broadens the receptive field while preserving feature map spatial dimension [17]. DeepLab employs Atrous Spatial Pyramid Pooling (ASPP) to aid in handling multi-scale images, controlling feature response density to obtain multi-scale context [17]. Resultantly, while FCNs and U-Net are more commonly used in biomedical image segmentation, DeepLab provides a deeper model architecture with a greater number of features, potentially better suited to complex segmentation tasks [31].

1.3.7. Generative adversarial networks (GAN)

GAN proposed by Goodfellow et al. encompass a variation of generative models specialised for the synthesis of images from real data [20,32]. GANs are composed of a generator and discriminator NN connected through back-propagation [20,33]. The generator network creates false images, and the discriminator is tasked at differentiating between fabricated and real images [20]. The discriminator network's loss reflects its misclassification rates and causes its weights to be updated through back-propagation [33]. Meanwhile, GAN's generator network utilises random input (random noise) to generate false images

[33].

In the context of image segmentation, replacing the generator network with a segmentation network enables the GAN to differentiate between predicted segmentation tasks and the ground truth [20]. However, this approach is associated with difficulties in training, and maintaining segmentation quality [34]. Resultantly, GAN variants have been developed, with one of the most successful being the segmentation adversarial network (SegAN), using a fully convolutional GAN for pixel-to-pixel segmentation [34].

A recent innovation that combines Swin Transformers with GANs is the Swin Transformer-based GAN for multi-modal medical image translation entitled MMTrans, coined by Yan et al. in 2022 [35]. MMTrans is composed of a generator based on the SwinIR architecture (Swin Transformer that can predict deformable vector fields), skilled at generating images within the same category of the modality of choice [35]. After the generator, there is a registration network that corrects any minor mismatches between source and target domain images [35]. Finally, MMTrans contains a discriminator, built using a CNN that discerns whether the target image is most like the generator or the real image [35].

1.3.8. Auto-Encoders (AE)

AE form an un-supervised variation of the feedforward NN (see in MLPs), characterised by an input layer with a feature number identical to the output layer [36]. Thus, the input layer is compressed into a "latent-space representation" and then reconstructed as an output [36]. Variational AEs (VAE) are fundamentally similar; however, they use actual samples to create an ideal distribution which is inputted into a decoder network to build generated samples [7]. If generated samples are close to actual samples, an AE is trained, and VAE is adjusted so that its encoder's output attunes to the target distribution, reducing the loss function [7].

In the context of image segmentation, Yu et al. suggested a modified VAE entitled the Segmentation AE (SAE), an un-supervised segmentation model that can use unlabelled inputs [37]. SAE's encoder intakes segmentation images that have been pre-trained on an anatomical atlas prior (spatial prior) [37]. The SAE network is then trained using a Gumbel-SoftMax relaxation enabling efficient parameter optimisation and eventual training through back-propagation [37]. VAE-GANs have also been suggested, encompassing an architecture that can generate realistic GAN-generated training images [7].

1.3.9. Recurrent neural networks (RNN)

CNNs classically approach segmentation on a pixel-by-pixel basis, conversely RNN can process a list or sequence of pixels at once, using information captured from the previous pixel to aid in the prediction of the subsequent one [20]. Thus, RNNs enable input from current or previous layers creating a "memory" [38]. The Gated Recurrent Unit (GRU) and Long-Short Term Memory (LSTM) models are two commonly used RNN variants [38]. Within cardiac image segmentation, RNNs are beneficial in imaging series such as cine CMR and Echo sequences, establishing connections between current and previous outputs. In addition, they are often combined with FCNs to optimise inter-slice knowledge and improve segmentation [20].

1.4. Advanced building blocks

Advanced building-blocks, employed as add-ons to core NN architectures, have been designed to enhance model robustness, efficiency and accuracy.

An adaptive convolutional kernel is a dynamic filter that can be applied to a convolutional layer, permitting changes to its weights that vary based on input image [39]. This filter undergoes a second convolution over the input image enabling accurate classification while minimising and reducing memory demands [39]. Therefore, adaptive kernels achieve heightened generalisation compared to traditional CNNs as they dynamically extract more appropriate features depending on the respective input image [39]. Atrous (or dilated) convolutional kernels allow appreciation for global context and holistic feature learning, without minimising segmentation map resolution [40]. These convolutions add holes in between kernel elements resulting in "inflation" of the kernel [41]. A dilutional rate is also added as an additional parameter denoting kernel width [41].

ASPP is designed to capture wide image context in segmentation tasks through convolutional feature layers with filters that have various sampling rates and fields-of-view (Fig. 5) [42]. Residual connections are skip-connections that allow gradient flow directly through the network [20]. Dense connections concatenate the feature map of the current layer with outputs from the previous layer [20].

Attention Gates (AGs) offer a solution to the computationally expensive nature of traditional CNN models, aiming to use model parameters and intermediate feature maps more efficiently [43]. AGs enable automatic structural focus with minimal supervision, delineating the features most relevant to a specific task, and repressing less relevant features and regions [43]. Resultantly, AGs eradicate the need for external structural localisation without compromising prediction accuracy, simultaneously reducing the computational overload associated with CNNs [43]. Multiplicative and additive attention are the two existing types of AGs that can be embedded into any CNN architecture [43].

Deep Supervision Modules (DSV) generate multiple segmentation maps at all levels of resolution, transposed to build secondary segmentation maps [44]. This is accomplished by up-sampling the element-wise sum of adjacent resolution segmentation maps until the highest resolution is reached [44]. Resultantly, DSV improves the number of features that can be learnt and optimises model convergence [45].

Ensemble learning combines numerous trained models to make a prediction, where models "vote" on the most common outcome, resulting in higher-accuracy predictions [20]. Moreover, transfer learning involves deploying learning obtained from solving one task *S*, to another task *T,* often through initialising the new model, *T's* weights using the pre-trained weights from model *S* [20,46]. These approaches attempt to prevent overfitting without compromising image, and thus, segmentation quality.

1.5. Loss functions

When training a deep neural network, stochastic gradient descent is typically used as the optimisation function of choice, loss functions are employed to learn the target in a more accurate and efficient manner [47].

Cross-entropy loss function, defined as the "difference between two probability distributions for a random variable or set of events" is the most popular loss function in image classification and segmentation [47]. Cross-entropy can be used to summarise probability errors in pixel-wise segmentation [20]. Mean-dice loss is another widely employed segmentation-specific function, built as an adaptation to the dice coefficient, a value that calculates the similarity between two images [47]. Weighted cross-entropy and weighted dice-loss, form two variations of the loss functions, using weighted loss terms to overcome class imbalance, and include rare classes or objects [20]. Unified focal loss generalises both dice loss and cross-entropy loss to tackle class imbalance within data sets [48]. This function allows a single hyperparameter to be fine-tuned as opposed to the six hyperparameters associated with traditional focal loss functions, making it more efficient [48].

2. Methodology

A literature search is conducted on four databases specialised in biomedical and/or ML research: namely, *PubMed, Medrxiv, Arxiv* and *Papers with code*. Search queries have been built for each database

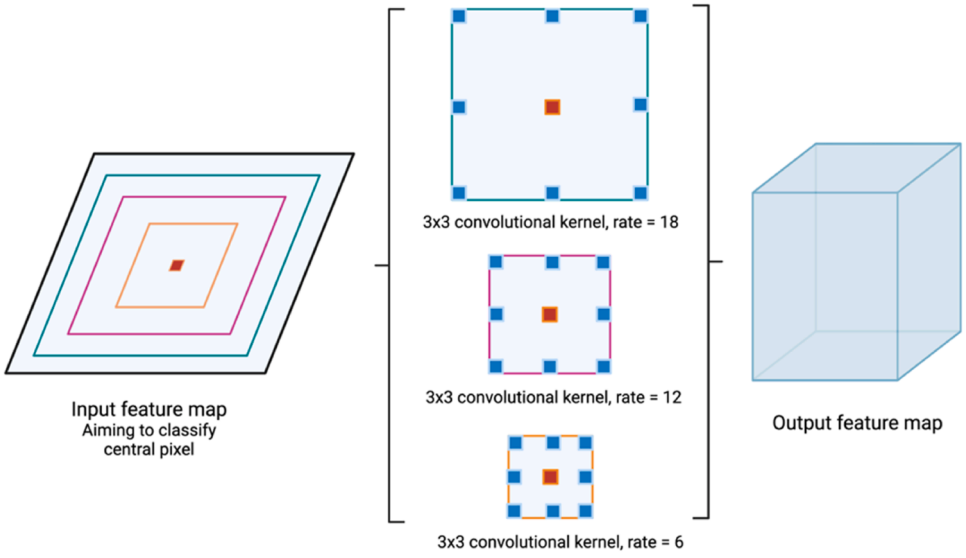

Fig. 5. Atrous spatial pyramid pooling aiming to classify the central, dark orange, pixel using 3 × 3 convolutional kernels with skip-connections.

aiming to identify studies meeting the pre-determined eligibility criteria (see Fig. 6 for search queries). As outlined in Fig. 4, a total of 277 papers resulted from the initial search, upon automatic and manual exclusion, 64 eligible studies were identified and included in the final data set.

2.1. Eligibility criteria

- Published online between January 1, 2019, and January 13, 2023.
- Article title directly refers to cardiac image segmentation.
- Article title OR Abstract outlines that a DL approach was utilised.
- Uses one or more of the following imaging modalities on human patients: CMR, CT, Echo, X-Ray.
- Primary Research Article.
- Written in English.
- Full text available.

We developed a framework to define the clinical and ML problem, outline the DL backbone of the suggested model, investigate the source and nature of training, testing and validation data, then evaluate the performance metrics, strengths, limitations, and areas of future investigation. This framework was a modification of the analytical framework proposed by Steven et al. publication, "Recommendations for Reporting Machine Learning Analyses in Clinical Research" [49].

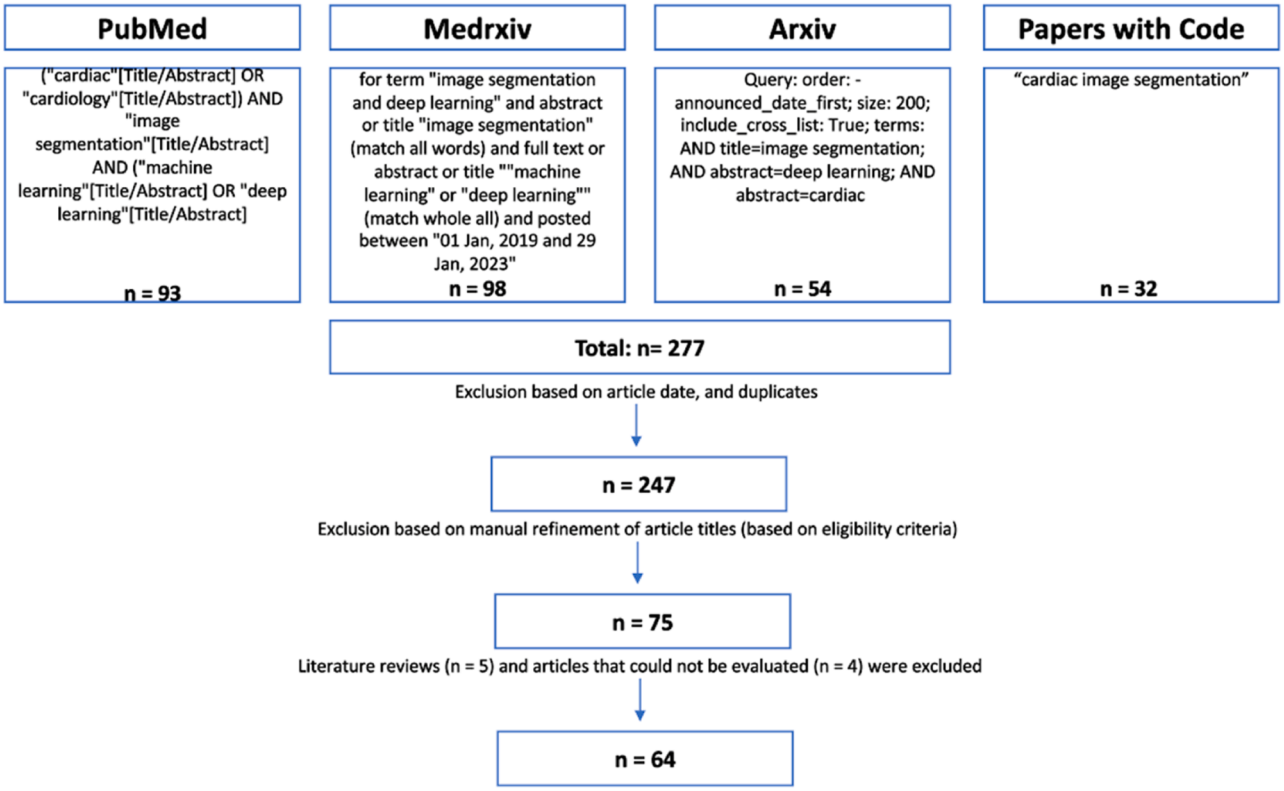

Fig. 6. Overview of the literature search containing queries used on each database, and the number of results following automatic and manual exclusion.

2.2. Proposed framework for analysis of DL papers in medical image segmentation

1 What is the clinical question?
2 What is the ML prediction problem?
3 Identify the backbone architecture.
  a Are any additional measures used?
4 Where does the data originate from?
  a Identify the source, nature, and quantity of the data.
  b Outline how training, validation and testing sets are split.
5 How is the proposed model evaluated?
6 What are the results of the model?
7 What are the strengths and limitations of the research?
8 What are potential areas of future investigation?

3. Results and discussion

3.1. Literature search results

The framework outlined in Section 2.2 is applied to analyse each study's proposed segmentation solution. The segmentation target, relative ranking, primary author, publication year, backbone architecture description, data source and split, and mean dice score are recorded, GitHub repositories are linked in the footnotes when available. The Mean Dice Score or Dice Similarity Coefficient (DSC) metric is the predominant evaluation metric employed to measure accuracy in image segmentation, representing the index of spatial overlap between two binary segments [50].

Tables 2, 3, 4, 5 and 6 present the results of CMR segmentation of the ventricles and myocardium, atria, pericardial adipose tissue, and the whole heart, respectively. Tables 7, 8, 9, 10 and 11, delineate segmentation results for CT-based segmentation of the ventricles and myocardium, atria, adipose tissue, aorta and coronary arteries, and whole heart, respectively. Tables 12, 13, and 14 demonstrate segmentation results from Echo, X-Ray Angiography, and multi-modal inputs, respectively. Studies are presented from highest to lowest mean DSC for each segmentation target, highlighted in the "Ranking" column.

3.2. Backbone architectures

U-Net was by far the most popular backbone architecture ($n = 44$), followed by CNN ($n = 9$) and FCN ($n = 6$), the frequencies of backbone architectures are expressed in Fig. 7. Of the four top-performing models presented in the subsequent section, three utilise a U-Net backbone, while one employs a nnU-Net foundation, representing a U-Net variant.

3.3. Datasets

Several segmentation challenges with unique datasets have taken place, with The Automated Cardiac Diagnosis Challenge (ACDC) and The Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) being two of the most prominent. Comprehending the nature of the data and top performers across these challenges enables a standardised comparison of segmentation architectures. In addition, of the 64 included studies in the review, approximately 23% ($n = 15$) utilised ACDC as their primary or supplementary dataset, while around 11% ($n = 7$) employed the M&Ms dataset.

3.3.1. ACDC

The Automated Cardiac Diagnosis Challenge (ACDC) comprises CMR images obtained from 150 patients at the University Hospital of Dijon obtained over a six-year period from two MRI scanners with magnetic strengths of 1.5T and 3.0T [111]. Patients were divided into five equally sized subgroups ($n = 30$ patients), namely, normal, previous myocardial infarction (LVEF $<$ 40%), dilated cardiomyopathy (LV volume $<$ 100 mL/m^2 and LVEF $<$ 40%), hypertrophic CM (LV cardiac mass $>$ 110 g/m^2, ventricular thickness $>$ 15 mm in diastole, and normal EF), and abnormal RV function (RV volume $>$ 110 mL/m^2, or RVEF $<$ 40%). Each patient's entry was accompanied with information regarding their weight, height, and diastolic/systolic phase instants.

This dataset was used in the MICCAI 2017 Conference in a challenge terminating in 2022. The training dataset was comprised of 100 patients (66.67%), 20 patients from each group, while testing data consisted of 50 patients (33.33%), 10 patients from each group. Ground-truth segments were developed by two experienced cardiologists. The segmentation targets were LV, RV and Myo. Results and architecture details of the three top-performing segmentation models from the ACDC challenge are presented in Table 15. Isensee's submission achieved first place across all three segmentation structures, utilising interconnected 2D and 3D U-Net models applied to each component of the cardiac cycle [112]. Similarly, Zotti's runner-up architecture employed a U-Net backbone coupled with cardiac shape-prior to augment performance [113]. In contrast, Painchaud's high-achieving third place architecture used an AE, aiming to alter anatomically implausible segmentation results into realistic outcomes, without affecting geometric or clinical metrics [114].

3.3.2. M&Ms

The Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) encompasses a data set used in the MICCAI 2020 Conference [115]. This dataset is composed of 375 CMR datasets obtained from four MRI scanners across six hospitals in three countries. Thus, when compared to previous challenges such as ACDC, M&Ms provides a heterogeneous compilation of data aiming to reflect the high degree of variability between images obtained from different vendors and locations. Patients included demonstrated diverse cardiac pathologies including hypertrophic cardiomyopathy, dilated cardiomyopathy, coronary heart disease, abnormal RV, myocarditis, ischaemic cardiomyopathy, and healthy volunteer.

The dataset was divided into 175 training cases (46.67%), 40 validation cases (10.67%), and 160 testing cases (42.66%). The top-three performing models within this dataset all employed a nnU-Net backbone architecture in addition to various data augmentation techniques (see Table 16). Utilising data augmentation methods such as parameter variation and intensity transformation helped build new training images, ultimately improving model generalisability. However, domain adaptation, typically combined with U-Net backbones, did not yield results superior to nnU-Net models without domain adaptation.

3.4. Top-performing segmentation models

The details of the top-performing models (achieving the highest mean DSC) in CMR ventricular and myocardial segmentation, CMR atrial segmentation, CT atrial and ventricular segmentation and CT aortic segmentation are presented in this section.

Hasan et al.'s CondenseU-Net is a top-performing CMR segmentation model achieving mean dice scores of 96.8%, 93.5% and 90.1% in LV, RV and Myo segmentation respectively [67]. This modified DenseNet combines CondenseNet and U-Net, replacing standard and group-convolutions with learned group-convolutions, enabling more relevant feature selection without increasing computational load [67]. Standard convolutions necessitate high computational power, while group-convolutions require predefined filters limiting representational abilities [67]. In contrast, the proposed learned group-convolutions employ multi-stage schemes that can dynamically learn representations during training [67]. CondenseU-Net is composed of two stages, the multi-condensing stage followed by the optimisation stage [67]. The former involves computing and averaging weights for every feature, then extracting a low-magnitude weighted column from the features [67]. The latter comprises the training phase, utilising a group-lasso regulariser, preventing over-fitting. While encouraging group-level sparsity by defining outgoing connections deriving from a single neuron as zero or non-zero [67]. Furthermore CondenseU-Net embraces

Table 2
Results of CMR ventricular and myocardial segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LV	1	Upendra et al. [51]	2019	2D U-Net model (a modified U-Net based off current state-of-the-art LV segmentation) + segmentation adversarial network (SegAN)	ACDC 2017	53% Training ($n = 80$) 13% Validation ($n = 20$) 33% Testing ($n = 50$)	95.22%
LV	2	Yan, Z. et al. [52]	2022	Improved SegNet: deep separable CNNs (Convoluted Neural Networks) + pyramid pooling + enhanced coder	1354 CMR scans	N/A	87.80%
LV	3	Wang, Zi Hao et al. [53]	2020	ResNet18 input + Dense FCN Output + ResNet34 for control point regression	HVC + ACDC	80% Training 10% Validation 10% Testing	86.40%
LV	4	Xiong, J. et al. [54]	2021	Deep RL + First P-Net (CNN) + Next-P-Net (point-centric concatenated matrix) + Deep Q Network	ACDC 2017 + Sunnybrook 2009	67% Training ($n = 10$) 33% Validation ($n = 5$)	N/A
LV	5	Wang et al. [55]	2021	Dense RNN with LSTMs	137 patients	95% Training ($n = 130$) 5% Testing ($n = 7$)	N/A
RV	1	Jabbar et al. [56]	2021	SA-LA model: multi-encoder-decoder U-Net + spatial context + deep supervision	M&Ms 2021	44% Training ($n = 160$) 11% Validation ($n = 40$) 44% Testing ($n = 160$)	90.30%
RV	2	Tran, C. T. et al. [57]	2020	2D U-Net	100 patients	50% Training ($n = 50$) 10% Validation ($n = 10$) 40% Testing ($n = 40$)	90.00%
LV+RV	1	Duan, J. et al. [58]	2019	2.5D FCN + anatomical shape prior knowledge	UK Digit Heart Project and Pulmonary Hypertension Dataset	UK Digital Heart Project: 55% Training ($n = 1000$) 45% Testing ($n = 831$) Pulmonary Hypertension: 68% Training ($n = 429$) 32% Testing ($n = 200$)	87.98%
LV+RV	2	Shen, D. et al. [59]	2021	3D Dense U-Net[1]	150 patients	76% Training ($n = 114$) 24% Testing ($n = 36$)	74.50%
Myocardial (Myo) scarring	1	Ding et al. [60]	2022	Siamese U-Net + Transformer Network (pre-LN transformer)	EMIDEC	Dataset divided into 5 parts, 4 used for training and validation, 1 used for testing	84.33%
Myo scarring + oedema	2	Wang, Kai-Ni et al. [61]	2022	AWSNet, Cascaded Anatomical Segmentation Network (vanilla U-Net) + Deep Auto-weighted Supervision Attention Network (ASAN) + Deep auto-weighted supervision (DAS) + Pixel-wise attention module[2]	MyoPS 2020	51% Training ($n = 23$) 4% Validation ($n = 2$) 44% Testing ($n = 20$)	70.65%
Myo scarring	3	Fahmy et al. [62]	2021	2D U-Net	191 patients	50% Training ($n = 81$) 25% Validation ($n = 40$) 25% Testing ($n = 41$) External Testing ($n = 29$)	N/A
LV+Scar	N/A	Popescu, D. M. et al. [63]	2022	Anatomical Convolutional Segmentation Network (ACSNet) U-Net with residuals + ResU-Net + style transfer for data enhancement	155 LGE-CMR Images + 246 synthetic "LGE-like" scans	Training (2484 images from two CMR sources) Testing (269 images from LGE-CMR set)	75.00%
LV+Myo	1	Liu, Yashu et al. [64]	2019	Res-U-Net + histogram matching[3]	MS-CMR 2019	Training: 35 fake images + 4 real images	93.15%
LV+Myo	2	Ankenbrand, M. J. et al. [65]	2021	U-Net + ResNet34 + transfer learning [4]	Kaggle	64% Training ($n = 14$) 23% Validation ($n = 5$) 13% Testing ($n = 3$)	84.99%
LV+Myo+RV	1	Oksuz et al. [66]	2020	U-Net + RNN (artefact detection network)	UK BioBank	75% Testing ($n = 3000$) 13% Validation ($n = 500$) 13% Testing ($n = 500$)	94.60%
LV+Myo+RV	2	Hasan et al. [67]	2020	U-Net + DenseNet (CondenseU-Net)	ACDC 2017	70% Training 15% Validation 15% Testing	93.45%
LV+Myo+RV	3	Zhang, J. et al. [68]	2022	Nested Capsule Dense Network (NCDN): FC-DenseNet with dense net replaced by a capsule dense block (CNN-based) [5]	ACDC	70% Training 10% Validation 20% Testing	91.77%
LV+Myo+RV	4	Fu, Z. et al. [69]	2022	TF-U-Net: Transformer + U-Net	ACDC + Synapse	70% Training ($n = 70$) 10% Validation ($n = 10$) 20% Testing ($n = 20$)	91.72%
LV+Myo+RV	5	Koehler, Sven et al. [70]	2020	U-Net + data augmentation	ACDC 2017 + GCN	ACDC Data: 75% Training ($n = 75$) 25% Testing ($n = 25$)	91.67%

(continued on next page)

Table 2 (*continued*)

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
						GCN Data: 75% Training ($n = 152$) 25% Testing ($n = 51$)	
LV+Myo+RV	6	Amirrajab, S. et al. [71]	2022	Generative Adversarial Network (GAN): U-Net with U-Net generator (image synthesis network)	M&Ms + synthetic data	67% Training ($n = 100$) 33% Testing ($n = 50$)	91.40%
LV+Myo+RV	7	Wibowo, A. et al. [72]	2022	2D U-Net + few-shot learning [6]	ACDC 2017	50% Training ($n = 50$) 50% Testing ($n = 50$)	90.89%
LV+Myo+RV	8	Upendra, R. et al. [73]	2020	SegAN model (segmentor network with an encoder-decoder FCN and critic network as the encoder)	ACDC 2017	80% Training 20% Validation	90.31%
LV+Myo+RV	9	Guo, F. et al. [74]	2020	CNN-guided kernel cut segmentation model	UK BioBank + ACDC	50% Training ($n = 50$) 10% Validation ($n = 10$) 40% Testing ($n = 40$)	90.07%
LV+Myo+RV	10	Zhang, Yao et al. [75]	2020	3D U-Net using label propagation + style transfer [7]	M&Ms	48% Training ($n = 185$) 52% Testing ($n = 200$)	89.22%
LV+Myo+RV	11	Campello, Víctor M. et al. [76]	2019	CNN + CycleGAN (image synthesis)[8]	MS-CMRSeg	42% Training ($n = 36$) 11% Validation ($n = 9$) 47% Testing ($n = 40$)	89.20%
LV+Myo+RV	12	Ma, Jun et al. [77]	2020	2D + 3D nnU-Net + Histogram-matching [9]	M&Ms	27% Training 20% Validation (For validation and training $n = 175$) 53% Testing ($n = 200$)	87.35%
LV+Myo+RV	13	Chen, D. et al. [11]	2022	Multi-Image Type Bidimensional U-Net (MI-U-Net)	72 patients	69% Training ($n = 50$) 9% Validation ($n = 7$) 21% Testing ($n = 15$)	86.67%
LV+Myo+RV	14	Vesal, S. et al. [78]	2021	Unsupervised Domain Adaptation modified DR-U-Net, based on Adversarial learning + entropy minimisation + output feature space alignment + point-cloud shape adaptation [10]	STACOM MS-CMRSeg 2019 and MM-WHS 2017 (cross-modality)	80% Training ($n = 16$) 20% Testing ($n = 4$)	86.00%
LV+Myo+RV	15	Scannell et al. [79]	2020	2D U-Net + domain adversarial learning[11]	M&Ms	47% Training ($n = 175$) 11% Validation ($n = 40$) 42% Testing ($n = 160$)	85.33%
LV+Myo+RV	16	Chen, C. et al. [80]	2020	2D U-Net [12]	UK BioBank + ACDC	87% Training ($n = 3975$) 13% Testing ($n = 600$)	84.33%
LV+Myo+RV	17	Chen, Xiang et al. [81]	2022	3D U-Net + co-attention block + Histogram Matching[13]	ACDC + M&M + UK BioBank	75% Training ($n = 1080$) 11% Validation ($n = 157$) 14% Testing ($n = 200$)	76.13%
LV+Myo+RV	18	Wang, H. et al. [82]	2022	Alternating Union Network (AUN) composed of ISN + LSN subnetworks	MS-CMRSeg 2019	76% Training ($n = 34$) 6% Validation ($n = 3$) 18% Testing ($n = 8$)	74.17%
LV+Myo+RV	19	Ma, Wanqin et al. [83]	2022	ResNet101 + Fourier transformations (cross-domain learning) + pseudo-labels	CMRxMotion	55% Training + Validation ($n = 200$) 45% Testing ($n = 160$)	73.77%
LV+Myo+RV	20	Song, L. et al. [84]	2021	Lightweight cross-consistency network (LCC—Net): U-Net + Ghost module	ACDC	80% Training ($n = 80$) 20% Testing ($n = 20$)	71.03%
LV+Myo+RV	21	Chen, Jingkun et al. [85]	2019	Adversarial segmentation network using DR-U-Net + discriminator model using CNN [14]	MICCAI 2019	N/A	69.10%
LV+Myo+RV	22	Gu et al. [86]	2022	Few-Shot Unsupervised Domain Adaptation (FUDA) made of Dilated-Residual U-NET (DR-U-NET) + target image generation (random adaptive instance normalisation) [15]	MS-CMRSeg 2019	N/A	62.67%

[1] https://github.com/dsc936/DenseUnet_for_TPM_segmentation.
[2] https://github.com/soleilssss/AWSnet/tree/master.
[3] https://github.com/Suiiyu/MS-CMR2019.
[4] https://github.com/chfc-cmi/cmr-seg-tl.
[5] https://github.com/jk1008611/NCDN.
[6] https://github.com/bowoadi/cine_MRI_segmentation_classification.
[7] https://github.com/YaoZhang93/Semi-supervised-Cardiac-Image-Segmentation-via-Label-Propagation-and-Style-Transfer.
[8] CycleGAN method: https://github.com/junyanz/pytorch-CycleGAN-and-pix.
[9] https://github.com/JunMa11/HM_DataAug.
[10] https://github.com/sulaimanvesal/PointCloudUDA.
[11] https://github.com/cianmscannell/da_cmr.
[12] https://github.com/cherise215/CardiacMRSegmentation.
[13] https://github.com/cistib/DDIR.
[14] https://github.com/jingkunchen/MS-CMR_miccai_2019.
[15] https://github.com/MingxuanGu/.

Table 3
Results of CMR atrial segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
Left Atrial (LA)	1	Xiong, Z. et al. [87]	2021	Double 3D-U-Net-based CNN	154 patients	65% Training ($n = 100$) 35% Testing ($n = 54$)	93.20%
LA	2	Uslu, F. et al. [88]	2022	TMS-Net (multi-view network)[1]	STACOM 2013 + 2018	70% Training ($n = 70$) 10% Validation ($n = 10$) 20% Testing ($n = 20$)	92.00%
LA	3	Hasan, S. M. K. et al. [89]	2021	Multi-Task Cross-Task Learning (MTCTL) using V-Net [2]	MICCAI STACOM 2018	80% Training ($n = 80$) 20% Validation ($n = 20$)	91.80%
LA + Right Atrial (RA)	N/A	Wang, Y., et al. [90]	2022	UU-Net + ResNet	150 patients	80% Training 20% Testing	96.70%
LA + Scar	N/A	Yang, G. et al. [91]	2020	Multiview two-task (MVTT) recursive attention model: Full CN (Caseous Necrosis) + dilated residual network + dilated attention network	190 patients	90% Training ($n = 153$) 10% Testing ($n = 17$)	90.00%

[1] https://github.com/fzehrauslu/TMS-Net.
[2] https://github.com/smkamrulhasan/MTCTL.

Table 4
Results of CMR aortic segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
Abdominal Aorta (AA)	1	Ruijsink, B. et al. [92]	2020	SemiQSeg: FCN + Quality control LSTM + Dense convolution	740 patients	86% Training ($n = 638$) 14% Testing ($n = 102$)	95.56%
AA	2	Chen, W. et al. [93]	2022	U-Net + attention module (XR-MSF-U-Net)	1204 CT images and 1345 CMR images	N/A	94.38%
AA	3	Li, Y. et al. [94]	2022	Self-attention mechanism + ESA-U-Net	150 patients	67% Training ($n = 100$) 33% Testing ($n = 50$)	91.50%
AA+ Coronary Arteries (CA)	4	Cheung, Wing Keung, et al. [95]	2021	2D automated modified U-Net	69 patients	64% Training ($n = 44$) 16% Validation ($n = 11$) 20% Testing ($n = 14$)	91.20%

Table 5
Results of CMR pericardial adipose tissue segmentation model outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score, training data details and github repository links.

Segmentation Target	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
Pericardial Adipose Tissue	Li, Zhuoyu et al. [96]	2022	PAT-CNN: 3D Res-U-Net-CNN	391 patients	86% Training + Validation 14% Testing	74.00%

Table 6
Results of CMR whole heart segmentation model outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
Whole Heart	Zhao, L. et al. [17]	2022	nn-TransU-Net [1]	ACDC + MSD + MyoPS	72% Training ($n = 144$) 18% Validation ($n = 36$) 10% Testing ($n = 20$)	93.60%

[1] Data augmentation method: https://github.com/MIC-DKFZ/batchgenerators.

Table 7
Results of CT ventricular and myocardial segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LV+LA+RV+RA+Myo	1	Bui et al. [97]	2022	U-Net + computer-generated labels + multi-atlas segmentation	1124 scans	76% Training (n = 851) 19% Validation (n = 213) 5% Testing (n = 60)	94.40%
RV	2	Zhao, Chen et al. [98]	2021	Spatial Temporal V-Net (ST-V-Net): V-Net + convolutional LSTM [1]	45 patients	N/A	85.36%
LV+LA+RV+RA+Myo+AA+ Pulmonary Vein (PV)	3	Zhao, Ziyuan et al. [99]	2022	Multi-scale multi-view global-local contrastive learning (MMGL): U-Net	MM-WHS 2017	2:1:1 Training: Validation: Testing	84.90%
Myo	4	Huang, Ziyi et al. [100]	2020	ReLayNet + Dropout-based Monte Carlo sampling	15 patients	N/A	60.50%

[1] https://github.com/MIILab-MTU/RV_segmentation.

Table 8
Results of CT atrial segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LA	Abdulkareem, M. et al. [101]	2022	U-Net [1]	337 patients	70% Training 15% Validation 15% Testing	88.50%

[1] https://github.com/mabdulkareem/lav_volume_with_qc.

Table 9
Results of CT adipose tissue segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
Epicardial fat volume	1	Siriapisith, T. et al. [45]	2021	3D U-Net with AG and DSV (AG-DSV-U-Net)	220 patients	73% Training (n = 160) 18% Validation (n = 40) 9% Testing (n = 20)	90.06%
Cardiac adipose tissue	2	Huang, Ziyi et al. [102]	2022	U-Net segmentation network + CAM-based pseudo-label generation	44 patients	N/A	88.38%

Table 10
Results of CT aortic and coronary artery segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
AA	1	Li, F. et al. [103]	2022	No-New-Net: nnU-Net	130 patients	68% Training (n = 88) 17% Validation (n = 22) 15% Testing (n = 20)	97.00%
AA	2	Chen, W. et al. [93]	2022	U-Net + attention module (XR-MSF-U-Net)	1204 CT images and 1345 CMR images	N/A	94.56%
AA+CA	3	Cheung, Wing Keung et al. [95]	2021	2D automated modified U-Net	69 patients	64% Training (n = 44) 16% Validation (n = 11) 20% Testing (n = 14)	91.20%

a bottleneck block and up-sampling pathway, creating memory-efficient, dense connections that allow selected features to be reused by multiple groups [67]. Overall, the proposed model enables accurate segmentation while minimising the number of parameters requiring training [67].

Wang et al.'s UU-NET proposes an CMR atrial segmentation model achieving a mean dice score of 96.70% for LA and RA segmentation [90]. The outlined approach suggests an improved U-Net, characterised by U-shaped upper and lower sampling layers, built using residual theory (ResNet) as the selected encoder-decoder. The suggested residual module aims to limit model depth by delaying gradient convergence during network propagation [90]. Furthermore, sampling modules aid

Table 11
Results of CT whole heart segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LV+LA+RV+RA+ Myo+AA+ Pulmonary Arteries (PA)	1	Yoshida, A. et al. [104]	2022	U-Net	20 patients	N/A	95.00%
LV+LA+RV+RA+ Myo+AA+PA	2	Park, S. et al. [105]	2022	U-Net + distance transformation	MM-WHS 2017	33% Training + Validation ($n = 20$) 67% Testing ($n = 40$)	87.00%

Table 12
Results of echo segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Author	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LV	1	Jafari, M. H. et al. [106]	2019	U-Net segmentation network + CNN critic network	427 patients	70% Training 10% Validation 20% Testing	92.00%
LV	2	Zhu et al. [107]	2021	U-Net + Active Contour (AC)	1500 patients	99% Training ($n = 1490$) 1% Validation ($n = 10$)	85.85%

Table 13
Results of X-ray angiography segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
CA	Iyer, Kritika et al. [31]	2021	CNN combining Angiographic processing DeepLab network (AngioNet) with a semantic segmentation network [1]	UM dataset + MM QCA data	Training + Validation: UM dataset separated into 5 partitions, 4 used for training and 1 used for validation Testing: 6th partition was used for testing	86.40%
CA	Zhu et al. [108]	2021	FCN + pretrained on ResNet + Dikted Network (PSPNet)	109 patients	2:1 Testing: Training	N/A

[1] https://github.com/kritiyer/AngioNet.

Table 14
Results of multi-modal (CT and CMR) segmentation models outlining: author, publication year, architecture description, data source and split, segmentation structure, mean dice score and training data details.

Segmentation Target	Ranking	Authors	Publication Year	Architecture Description	Data Source	Data Split	Mean Dice Score
LV	1	Huang, Xiaoqiong et al. [109]	2020	2D U-Net + zero-shot style transfer	M&Ms	60% Training ($n = 150$) 40% Testing ($n = 100$)	89.72%
LV+Myo+RV	2	Chartsias, A. et al. [110]	2019	FCN (spatial decomposition network) [1]	ACDC 2017 + 26 patients + MM-WHS 2017 + 10 canines	70% Training 15% Validation 15% Testing	83.90%
AA+LA+LV+Myo	3	Vesal, S. et al. [78]	2021	Unsupervised Domain Adaptation modified DR-U-Net, based on Adversarial learning + entropy minimisation + output feature space alignment + point-cloud shape adaptation[2]	STACOM MS-CMRSeg 2019 and MM-WHS 2017 (cross-modality)	80% Training ($n = 16$) 20% Testing ($n = 4$)	72.50%

[1] https://github.com/agis85/anatomy_modality_decomposition.
[2] https://github.com/sulaimanvesal/PointCloudUDA.

in maintaining accuracy when increasing feature complexity, by connecting the prior U-Net decoder with the subsequent U-Net encoder [90]. Additionally, this module builds numerous paths for data transmission, utilising features of FCNs in the U-Net paths [90]. In addition, deep deconvolutions are incorporated into training and testing stages to provide a supervised learning method [90]. Therefore, the combination of augmented complexity and various connected pathways creates an accurate and efficient atrial segmentation model.

Bui et al.'s DeepHeartCT proposes a high-performing fully automated framework for CT coronary angiography image segmentation, achieving DSCs of 98.0%, 94.0%, 94.0%, 93.0% 93.0% in LV, LA, RV, RA, and Myo segmentation, respectively. This U-Net-based architecture is fortified by combined multi-atlas and corrective segmentation (CMACs) [97]. CMACs uses bounding box detection to determine the

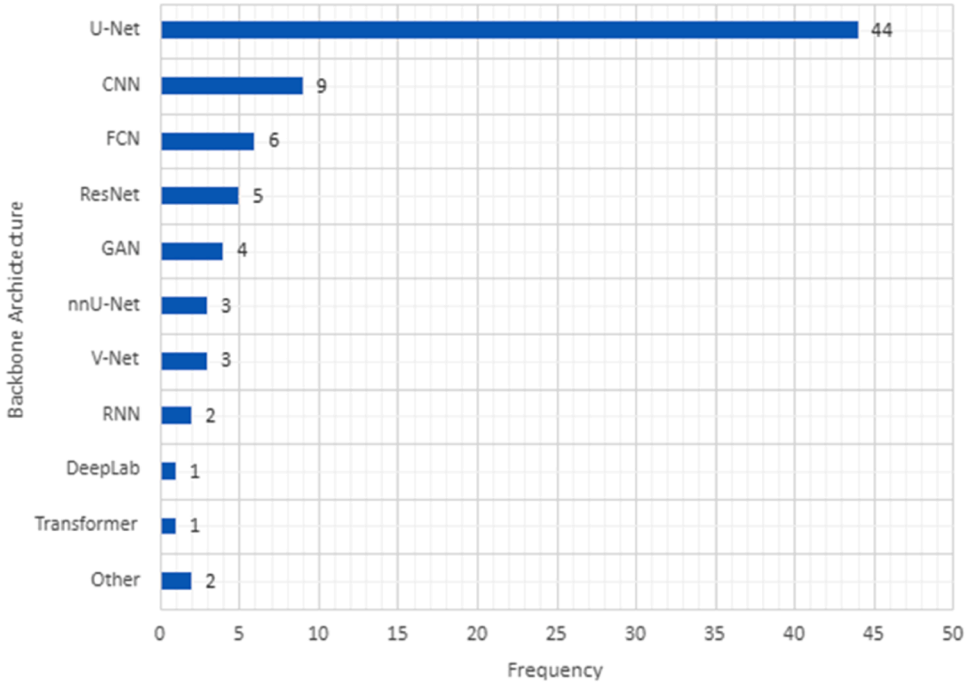

Fig. 7. Frequency of backbone architectures employed by each paper included in the review.

Table 15
Results of top-three performing models on the ACDC dataset including author, publication year, architecture description, segmentation structure and mean dice score.

Authors	Publication Year	Architecture Description	Segmentation Target	Mean Dice Score
Isensee et al. [112]	2018	Modified interconnected 2D and 3D U-Net + Batch normalization + leaky ReLU nonlinearity + Deep supervision	LV RV Myo	94.75% 92.50% 90.75%
Zotti et al. [113]	2019	U-Net (multi-resolution grid architecture) + cardiac shape-prior	LV RV Myo	93.80% 90.95% 89.40%
Painchaud et al. [114]	2019	Adversarial variational autoencoder + anatomically constrained data augmentation	LV RV Myo	93.60% 90.85% 88.90%

Table 16
Results of top-three performing models on the M&M dataset including author, publication year, architecture description, segmentation structure and mean dice score.

Authors	Publication Year	Architecture Description	Segmentation Target	Mean Dice Score
Full et al. [115]	2020	nnU-Net + data augmentation (gaussian-noise, brightness, gamma)	LV RV Myo	91.25% 88.50% 85.30%
Zhang et al. [115]	2020	nnU-Net + data augmentation (histogram matching, gamma)	LV RV Myo	90.90% 87.95% 84.55%
Ma et al. [115]	2020	nnU-Net + data augmentation (histogram matching, gamma)	LV RV Myo	90.50% 87.50% 84.05%

image region encompassing the whole heart. This is coupled with multi-atlas segmentation and label generation to create an annotated training dataset. Following this, corrective segmentation alters the labels generated in the preceding step to differentiate cardiac tissue from other intra-thoracic structures [97]. Reverse ranking is used to assess the quality of the computer-generated labels obtained through CMACs, and the selected labelled images are inputted into the U-Net segmentation model. Therefore, the model is trained on computer-generated examples, but is validated using manually annotated, real labels [97]. DeepHeartCT can segment up to 12 structures simultaneously, utilising a dice loss function and ReLu activation function after each layer [97]. Thus, DeepHeartCT successfully overcomes obstacles to adequately annotated and high-quality training data, without compromising training time or segmentation accuracy. nnU-Net is a fully automated aorta, aortic valve and LV outflow tract segmentation model achieving mean DSCs of 97% using cardiac CT images [103]. This study is the first to employ nnU-Net for cardiac image segmentation, providing the pivotal benefit of automatic input data pre-processing, and parameter and hyperparameter fine-tuning within a U-Net structure [103]. Thus, nnU-Net can augment training data through image cropping, resampling, and data normalisation to reduce artefact [103]. In addition, No-New-Net provides the additional benefit of easy conversion to stereolithography files required for 3D evaluations prior to trans-catheter aortic valve interventions, giving it a direct clinical advantage [103].

3.5. Key challenges in cardiac image segmentation

There are a variety of key obstacles limiting the clinical application of high-performing segmentation models, or conversely restricting the accuracy of more robust models.

3.5.1. Inadequate data availability and data quality

Firstly, the stark contrast between high task complexity and limited availability of adequate quality, labelled, training images remains to be

an obstacle when building deep NNs in cardiac image segmentation. As a result, models are prone to over-fitting, and efficient classification requires the deployment of further advanced techniques. For example, two regularisation techniques, weight regularisation and dropout, are often used to optimise learning. The former involves adding weight penalties to the loss function based on the relevance of the input, while the latter "drops" random components from the NN during training, driving sparse representation [20].

As an attempt to increase the magnitude of training data, strategies such as cross-modality image segmentation have been proposed. This method uses feature adaptation to alter an input image from an undesired imaging modality to the modality of choice, indicating that CT images could be used to train a CMR segmentation model [116]. Multi-atlas-based segmentation proposes an alternative approach where an anatomical atlas library containing pre-segmented cardiac structures is transformed into target images for the segmentation model [117].

The nature of medical imaging is dynamic and inextricably influenced by involuntary organ motion, patient movement or breathing, and challenges with image acquisition [84]. Resultantly, acquired images suffer from a variety of artefacts that can hinder segmentation, causing misleading and inaccurate results [84]. Image de-noising is a domain of medical image processing separate to segmentation, however, combining both approaches in one framework can help overcome challenges in image quality. Oksuz et al. suggests an end-to-end pipeline for artefact detection, correction, and segmentation [84]. This method reconstructs high quality CMRs (Cardia Magnetic Resonance) using a joint loss function, then leverages a data consistency term (k-space line detection network) to reconstruct under-sampled images [84]. Hence, proposing a framework to ensure high quality data despite inevitable motion artefacts [84].

3.5.2. Domain shift

Despite achieving exceptional segmentation accuracy on large, labelled data sets, cardiac segmentation models frequently experience plunges in performance on testing sets from heterogenous distributions [93]. In other words, even within one imaging modality, different imaging sites, vendors, and imaging protocols causes a domain shift leading to model underperformance [93]. Although data augmentation strategies may avoid overfitting to one domain, it does not directly improve model generalisability. Consequently, domain adaptation and generalisation have been introduced [93]. Huang et al. proposes a style-invariant cardiac segmentation model that utilises domain adaptation and generalisation through zero-shot style transfer [93]. This style transfer technique removes any appearance shift, creating a content image characterised by a low-level visual style that maintains semantic structure [93]. This segmentation model was tested on CMR images from four vendors, showing a 2% increase in DSC (Dice Similarity Coefficient) compared to baseline models in three, but a decline in performance on the fourth, thus requiring further investigation [93]. Vesal et al. propose an alternative unsupervised domain adaptation approach using adversarial learning [78]. The proposed framework integrates entropy minimisation, output space alignment, and shape-prior using point-cloud adaptation within a multi-task segmentation model [78]. This study successfully minimises cross-modality performance degradation, achieving the best DSC (87.3%) compared to benchmark models within multi-class segmentation [78].

3.5.3. Clinician mistrust

The clinical application and uptake of ML systems is often hindered by the "black box" phenomenon, where the prediction methodology is not directly decipherable or explainable [20]. Furthermore, evidence suggests that image-based medical DL systems can easily be biased by visually indiscernible alterations to input images leading to mis-segmentation [20]. As a result, quality-control methods have been proposed, such as Uslu et al.'s TMS-Net [88]. This method uses multi-view decoders to create high-quality segments despite varying levels of image noise, as well as "poor" segmentation masks from engineered noisy images to simulate under-training and poor image settings [88]. An unsupervised run-time quality estimation is then applied to distinguish between "good" and "bad" quality segments, aiming to reassure clinicians that inadequate quality segments can be self-detected [88]. The adoption of quality-control methods that can easily be understood and employed by clinicians may aid in increasing uptake, while minimising the effects of potential system error.

3.5.4. Standardised evaluation metrics

Within the domain of biomedical image segmentation, evaluation bias secondary to inaccurate metric employment or omission of "hold-out set sampling" necessary for validation indicates that model performance is not always accurately assessed [118]. In other words, published papers are prone to "cherry-picking" metrics to demonstrate accuracies approaching 100% [118]. Muller et al. suggests that evaluation bias represents a severe obstacle to widespread clinical applicability and proposes a guideline to evaluate research reliability in image segmentation [118]. To safely implement image segmentation models within clinical settings, doubts regarding evaluation bias and true model accuracy must be addressed.

3.6. Areas of future investigation

With the exponential progress, innovation and achievements made within the field of cardiac image segmentation, advancements to improve model efficiency and running time, cross-modality training, weak supervision, and end-to-end segmentation frameworks, amongst others, provide promising avenues for future research.

Swin transformers provide a stark benefit in comparison to classically employed FCNs or U-Net models as they enable a wider range of information to be captured without compromising running time [119]. Swin U-Net Transformers have achieved state-of-the-art results in brain tumour semantic segmentation [119]. However, when applied to cardiac segmentation, models such as Grzeszczyk et al.'s Multi-task Swin U-Net Transformer for motion artefact classification and CMR segmentation have only reached mean dice scores of 87.1% [120]. While this model provides a key advantage as it can classify motion artefact and perform image segmentation in one step, increasing training data and employing data augmentation techniques will improve segmentation accuracy [120].

Furthermore, image-to-image translation is a growing field within biomedical image segmentation, aiming to transform images obtained from one modality into images characteristic of another modality [35]. This provides evident benefits in terms of reducing the number of additional scans required, while improving diagnostic accuracy, and the complexity of image analysis. Swin Transformer-based GANs (Generative Adversarial Networks) have been used to facilitate multi-modal image translation, where Yan et al.'s study successfully translates CMR brain scans [35]. Implementing a similar architecture in cardiac imaging will significantly aid in overcoming access to sufficient high-quality training data, while reducing clinician workload and imaging-associated costs

To overcome limited access to large, annotated datasets weakly supervised, or unsupervised approaches to training are necessitated. Despite the proposition of attempts to reduce supervision such as few-shot learning and weakly-supervised learning, these methods have several disadvantages when implemented [121]. Few-shot learning, a semi-supervised data augmentation technique, is prone to noisy results that place too great emphasis on the available labelled data points [121]. In addition, weak supervision is not effective in leveraging the full potential of high-quality images [121]. Therefore, novel weakly supervised or unsupervised training approaches are required. Hooper et al. suggests a framework combining few-shot learning and weak supervision to overcome their respective limitations, however further modifications can fortify the performance of weakly supervised models

[121]. For example, introducing self-supervised pre-training modules able to gain global and local insight through domain and problem-specific cues using contrasting learning strategies [122]. Furthermore, building networks that can automate ROI and extreme point selection can help reduce the supervision-level required for successful model deployment [123].

As innovation within the biomedical imaging domain progresses, user-friendly, end-to-end image processing frameworks are becoming increasingly necessary. Medial Open Network for AI (MONAI) is an open source, easily operated, end-to-end biomedical imaging platform, enabling image labelling, transformation, segmentation, and model deployment [124]. Thus, MONAI aims to integrate state-of-the-art findings in biomedical imaging DL solutions into a single platform, driving scientific progression in the field [124].

3.7. Methodological limitations

Certain methodological modifications could expand the comprehensive nature of this literature review. For example, including a search on the publications defined by keywords rather than only relying on direct references to cardiac segmentation in the articles' title would broaden the scope of the eligibility criteria, identifying potentially missed articles. Similarly, including more databases such as Biorxiv and Google Scholar would diversify the included papers. Furthermore, increasing the selected timeline could ensure that several key papers, such as top performers in the MICCAI 2017 segmentation challenge are included.

4. Conclusion

In this paper, we provide an overview of the most pertinent neural network architectures, advanced building-blocks and loss functions within the field of cardiac image segmentation. In addition, we conduct a review of the current literature, exploring over 60 biomedical segmentation papers, proposing solutions to atrial, ventricular, myocardial, aortic, and coronary artery segmentation using images from CMR, CT, ultrasound, and X-ray.

Over the last decade there has been a steep rise within the field of cardiac ML. Initiatives such as euCanSHare (http://www.eucanshare.eu/) have led to the establishment of international, multi-cohort cardiovascular research platforms, driving innovation, and increasing the prospect of clinical application. Cardiac image segmentation has been fundamental to this, aiming to reach a highly personalised, patient-centred, accurate and time-efficient approach to the diagnosis and management of cardiac pathologies. Image segmentation forms a field of biomedical ML with high clinical acceptance, as it reduces clinician workload without significantly intervening in their decision-making process [5].

Despite the widespread popularity of U-Net-based architectures, their limited receptive fields, dependency on labelled data, tendency to over-fit and perform poorly on a test set, and susceptibility to poor image quality and artefact indicate that novel approaches are in need. As such, recent propositions including nnU-Net and Swin Transformers, alongside various data augmentation techniques such as image-to-image translation and end-to-end image processing will increase the clinical applicability, uptake and benefit of automated cardiac image segmentation.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Dr Aung recognises funding support from NIHR Integrated Academic Training Programme and the Academy of Medical Sciences.

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.

Data availability

No data was used for the research described in the article.

References

[1] Lekadir K, Leiner T, Young AA, Petersen SE. Editorial: current and future role of artificial intelligence in cardiac imaging. Front Cardiovasc Med 2020;7.
[2] Liu D, Jia Z, Jin M, Liu Q, Liao Z, Zhong J, et al. Cardiac magnetic resonance image segmentation based on convolutional neural network. Comput Methods Programs Biomed 2020;197:105755. Dec.
[3] Rehman R, Yelamanchili VS. Cardiac imaging [Internet]. National Center for Biotechnology Information. U.S. National Library of Medicine; 2023 [cited 2023Apr16]. Available from: https://pubmed.ncbi.nlm.nih.gov/28846331/.
[4] Larrey-Ruiz J, Morales-Sánchez J, Bastida-Jumilla MC, Menchón-Lara RM, Verdú-Monedero R, JL SG. Automatic image-based segmentation of the heart from CT scans. EURASIP J Image Video Process 2014;2014(1). https://doi.org/10.1186/1687-5281-2014-52. Nov 25.
[5] Valzania C, Gadler F, Maret E, Eriksson MJ. Cardiovascular imaging applications in clinical management of patients treated with cardiac resynchronization therapy. Hearts 2020;1(3):166–80. https://doi.org/10.3390/hearts1030017.
[6] Hassani C, Saremi F, Varghese BA, Duddalwar V. Myocardial radiomics in cardiac CMR. Am J Roentgenol 2020;214(3):536–45. Mar.
[7] Song Y, Ren S, Lu Y, Fu X, Wong KKL. Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. Comput Methods Programs Biomed 2022;220:106821. Jun.
[8] Zhou SK. Handbook of medical image computing and computer assisted intervention. In SMPA.R. Academic Press; 2020. p. 429–56.
[9] Goodfellow I, Bengio Y, Courville A. Introduction. Deep learning. MIT Press; 2016. p. 1–26.
[10] Seo B., Mariano D., Beckfield J., Madenur V. https://library.ucsd.edu/dc/object/bb47487939/_3_1.pdf. UCSD. 2019 Sep17.
[11] Chen D, Bhopalwala H, Dewaswala N, Arunachalam SP, Enayati M, Farahani NZ, et al. Deep neural network for cardiac magnetic resonance image segmentation. J Imaging 2022;8(5):149. May23.
[12] Some K. The history, evolution and growth of Deep Learning [Internet]. Anal Insight 2018. Analytics Insight[cited 2023Apr16]. Available from: https://www.analyticsinsight.net/the-history-evolution-and-growth-of-deep-learning/.
[13] Perceptron CAL. The artificial neuron (an essential upgrade to the McCulloch-Pitts Neuron) [Internet]. Medium. Towards Data Sci 2020 [cited 2023Apr16]. Available from: https://towardsdatascience.com/perceptron-the-artificial-neuron-4d8c70d5cc8d#:~:text=Frank%20Rosenblatt%2C%20an%20American%20psychologist,to%20as%20the%20perceptron%20model.
[14] Chandra AL. McCulloch-Pitts Neuron - mankind's first mathematical model of a biological neuron [Internet]. Medium. Towards Data Sci 2022 [cited 2023Apr16]. Available from: https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1.
[15] Introduction multilayer perceptron neural networks [Internet]. DTREG; 2023 [cited Apr16]. Available from: https://www.dtreg.com/solution/multilayer-perceptron-neural-networks.
[16] Multilayer perceptron [Internet]. DeepAI. DeepAI; 2019 [cited 2023Apr16]. Available from: https://deepai.org/machine-learning-glossary-and-terms/multilayer-perceptron.
[17] Chen L.C., Papandreou G., Schroff F. Rethinking atrous convolution for semantic image segmentation. 2017Jun.
[18] Mack C. Machine learning fundamentals (i): cost functions and gradient descent [Internet]. Medium. Towards Data Science; 2021 [cited 2023Apr16]. Available from: https://towardsdatascience.com/machine-learning-fundamentals-via-linear-regression-41a5d11f5220#:~:text=Put%20simply%2C%20a%20cost%20function,to%20as%20loss%20or%20error.
[19] Backpropagation [Internet]. DeepAI; 2020 [cited 2023Apr16]. Available from: https://deepai.org/machine-learning-glossary-and-terms/backpropagation.
[20] Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. Front Cardiovasc Med 2020;7.
[21] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Lect Notes Comput Sci 2015:234–41. Nov18.
[22] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2015. Mar8.
[23] Silver D. Literature review: fully Convolutional Networks [Internet]. Medium 2018. Medium[cited 2023Apr25]. Available from: https://medium.com/self-driving-cars/literature-review-fully-convolutional-networks-d0a11fe0a7aa.
[24] Srikrishna M, Heckemann RA, Pereira JB, Volpe G, Zettergren A, Kern S, et al. Comparison of two-dimensional- and three-dimensional-based U-Net

architectures for brain tissue classification in one-dimensional brain CT. Front Comput Neurosci 2022:15.
[25] Lin N, Lu H, Gao J, Qiao S, Li X. Vnet: a versatile network for efficient real-time semantic segmentation. In: Proceedings of the IEEE 37th international conference on computer design (ICCD); 2019. Nov10.
[26] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. NNU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203–11. Feb18.
[27] Zhao L, Zhou D, Jin X, NN-TransUNet ZW. An automatic deep learning pipeline for heart MRI segmentation. Life 2022;12(10):1570. Oct9.
[28] Dosovitskiy A, Beyer L, Kolesnikov A. An image is worth 16x16 words: transformers for image recognition at scale. In: Proceedings of the international conference on learning representations; 2023. 202ADSep28.
[29] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV); 2021. Aug17.
[30] Ahmed I, Ahmad M, Khan FA, Asif M. Comparison of deep-learning-based segmentation models: using top view person images. IEEE Access 2020;8:136361–73. Jul23.
[31] Iyer K., Najarian C.P., Fattah A.A., Arthurs C.J., Soroushmehr S.M.R., Subban V., et al. AngioNet: a convolutional neural network for vessel segmentation in X-ray angiography. 2021.
[32] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020;63(11):139–44.
[33] Overview of gan structure| machine learning | google developers [Internet]. Google. Google; 2023 [cited Apr25]. Available from: https://developers.google.com/machine-learning/gan/gan_structure.
[34] Decourt C, Duong L. Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI. Comput Biol Med 2020;123:103884. Aug.
[35] Yan S, Wang C, Chen W, Lyu J. Swin transformer-based gan for multi-modal medical image translation. Front Oncol 2022:12. Aug8.
[36] Dertat A. Applied deep learning - part 3: autoencoders [Internet]. Medium. Towards Data Sci 2017 [cited 2023Apr25]. Available from, https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798.
[37] Yu EM, Iglesias J, Dalca AV. An auto-encoder strategy for adaptive image segmentation. In: Proceedings of the machine learning research; 2020.
[38] Li J. Better cardiac image segmentation by highly recurrent neural networks [Internet]. University of California. UC San Diego; 2020 [cited 2023Apr25]. Available from: https://escholarship.org/uc/item/8dp4r0sp.
[39] Zamora Esquivel J, Cruz Vargas A, Lopez Meyer P, Tickoo O. Adaptive convolutional kernels. In: Proceedings of the IEEE/CVF international conference on computer vision workshop (ICCVW); 2019. Oct.
[40] Holländer B. U-Net, dilated convolutions and large convolution kernels in deep learning [Internet]. Medium 2018 [cited 2023Apr25]. Available from: https://medium.com/@branislav.hollander/U-Net-dilated-convolutions-and-large-convolution-kernels-in-deep-learning-a849f06ffb82.
[41] Yu F., Koltun V. Multi-scale context aggregation by dilated convolutions [Internet]. arXiv.org. 2016 [cited 2023Apr25]. Available from: https://arxiv.org/abs/1511.07122v3.
[42] Chen L.C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [Internet]. arXiv.org. 2017 [cited 2023Apr25]. Available from: https://arxiv.org/abs/1606.00915.
[43] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: learning to leverage salient regions in medical images. Med Image Anal 2019;53:197–207. Feb.
[44] Tureckova A., Turecek T., Kominkova Z., Rodŕıguez-Sánchez A. Kits challenge: vnet with attention gates and deep supervision. Submissions to the 2019 kidney tumor segmentation challenge: kits19. 2019.
[45] Siriapisith T, Kusakunniran W, Haddawy P. A 3D deep learning approach to epicardial fat segmentation in non-contrast and post-contrast cardiac CT images. PeerJ Comput Sci 2021;7.
[46] Karimi D, Warfield SK, Gholipour A. Transfer learning in medical image segmentation: new Insights from analysis of the dynamics of model parameters and learned representations. Artif Intell Med 2021;116:102078. Apr.
[47] Jadon S. A survey of loss functions for semantic segmentation [Internet]. 2020 arXiv.org. [cited 2023Apr25]. Available from: https://arxiv.org/abs/2006.14822.
[48] Yeung M, Sala E, Schönlieb CB, Rundo L. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput Med Imaging Graph 2022;95:102026. https://doi.org/10.1016/j.compmedimag.2021.102026. Jan.
[49] Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. Circ Cardiovasc Qual Outcomes 2020;13(10).
[50] Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. Acad Radiol 2004;11(2):178–89. Feb.
[51] Upendra RR, Dangi S, Linte CA. An adversarial network architecture using 2D U-Net models for segmentation of left ventricle from cine cardiac CMR. Funct Imaging Model Heart 2019;11504:415–24.
[52] Yan Z, Su Y, Sun H, Yu H, Ma W, Chi H, et al. SegNet-based left ventricular MRI segmentation for the diagnosis of cardiac hypertrophy and myocardial infarction. Comput Methods Programs Biomed 2022;227:107197.
[53] Wang Z., et al. Fully automated segmentation of the left ventricle in magnetic resonance images2020 July 01, 2020:[arXiv:2007.10665 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv200710665W.
[54] Xiong J, Po LM, Cheung KW, Xian P, Zhao Y, Rehman YAU, et al. Edge-sensitive left ventricle segmentation using deep reinforcement learning. Sensors 2021;21(7) (Basel).
[55] Wang Y, Zhang W. A dense RNN for sequential four-chamber view left ventricle wall segmentation and cardiac state estimation. Front Bioeng Biotechnol 2021;9:696227.
[56] Jabbar S., Talha Bukhari S., Mohy-ud-Din H. Multi-view SA-LA Net: a framework for simultaneous segmentation of RV on multi-view cardiac MR Images2021 October 01, 2021:[arXiv:2110.00682 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv211000682J.
[57] Tran CT, Halicek M, Dormer JD, Tandon A, Hussain T, Fei B. Fully automated segmentation of the right ventricle in patients with repaired Tetralogy of Fallot using U-Net. Proc SPIE Int Soc Opt Eng 2020:11317.
[58] Duan J, Bello G, Schlemper J, Bai W, Dawes TJW, Biffi C, et al. Automatic 3D Bi-ventricular segmentation of cardiac images by a shape-refined multi- task deep learning approach. IEEE Trans Med Imaging 2019;38(9):2151–64.
[59] Shen D, Pathrose A, Sarnari R, Blake A, Berhane H, Baraboo JJ, et al. Automated segmentation of biventricular contours in tissue phase mapping using deep learning. NMR Biomed 2021;34(12):e4606.
[60] Ding Y, Xie W, Wong KKL, Liao Z. DE-MRI myocardial fibrosis segmentation and classification model based on multi-scale self-supervision and transformer. Comput Methods Programs Biomed 2022;226:107049.
[61] Wang K.N., Yang X., Miao J., Li L., Yao J., Zhou P., et al. AWSnet: an auto-weighted supervision attention network for myocardial scar and edema segmentation in multi-sequence cardiac magnetic resonance images2022 January 01, 2022:[arXiv:2201.05344 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220105344W.
[62] Fahmy AS, Rowin EJ, Chan RH, Manning WJ, Maron MS, Nezafat R. Improved quantification of myocardium scar in late gadolinium enhancement images: deep learning based image fusion approach. J Magn Reson Imaging 2021;54(1):303–12.
[63] Popescu DM, Abramson HG, Yu R, Lai C, Shade JK, Wu KC, et al. Anatomically informed deep learning on contrast-enhanced cardiac magnetic resonance imaging for scar segmentation and clinical feature extraction. Cardiovasc Digit Health J 2022;3(1):2–13.
[64] Liu Y., Wang W., Wang K., Ye C., Luo G. An automatic cardiac segmentation framework based on multi-sequence MR Image2019 September 01, 2019:[arXiv:1909.05488 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv190905488L.
[65] Ankenbrand MJ, Lohr D, Schlötelburg W, Reiter T, Wech T, Schreiber LM. Deep learning-based cardiac cine segmentation: transfer learning application to 7T ultrahigh-field MRI. Magn Reson Med 2021;86(4):2179–91.
[66] Oksuz I, Clough JR, Ruijsink B, Anton EP, Bustin A, Cruz G, et al. Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation. IEEE Trans Med Imaging 2020;39(12):4001–10.
[67] Hasan SMK, Linte CA. CondenseUNet: a memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation. Proc SPIE Int Soc Opt Eng 2020:11315.
[68] Zhang J, Zhang Y, Zhang H, Zhang Q, Su W, Guo S, et al. Segmentation of biventricle in cardiac cine MRI via nested capsule dense network. PeerJ Comput Sci 2022;8:e1146.
[69] Fu Z, Zhang J, Luo R, Sun Y, Deng D, Xia L. TF-Unet:an automatic cardiac MRI image segmentation method. Math Biosci Eng 2022;19(5):5207–22.
[70] Koehler S, Tandon A, Hussain T, Latus H, Pickardt T, Sarikouch S, et al. How well do U-Net-based segmentation trained on adult cardiac magnetic resonance imaging data generalize to rare congenital heart diseases for surgical planning? Medical Imaging 2020: image-Guided Procedures. Robot Interv Model 2020. March 01, 2020.
[71] Amirrajab S, Al Khalil Y, Lorenz C, Weese J, Pluim J, Breeuwer M. Label-informed cardiac magnetic resonance image synthesis through conditional generative adversarial networks. Comput Med Imaging Graph 2022;101:102123.
[72] Wibowo A, Triadyaksa P, Sugiharto A, Sarwoko EA, Nugroho FA, Arai H, et al. Cardiac disease classification using two-dimensional thickness and few-shot learning based on magnetic resonance imaging image segmentation. J Imaging 2022;8(7).
[73] Upendra RR, Dangi S, Linte CA. Automated segmentation of cardiac chambers from cine cardiac MRI using an adversarial network architecture. Proc SPIE Int Soc Opt Eng 2020:11315.
[74] Guo F, Ng M, Goubran M, Petersen SE, Piechnik SK, Neubauer S, et al. Improving cardiac MRI convolutional neural network segmentation on small training datasets and dataset shift: a continuous kernel cut approach. Med Image Anal 2020;61:101636.
[75] Zhang Y., Yang J., Hou F., Liu Y., Wang Y., Tian J., et al. Semi-supervised cardiac image segmentation via label propagation and style transfer2020 December 01, 2020:[arXiv:2012.14785 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv201214785Z.
[76] Campello V.M., Martín-Isla C., Izquierdo C., Petersen S.E., González Ballester M. A., Lekadir K. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac MRI2019 September 01, 2019:[arXiv:1909.01182 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv190901182C.

[77] Ma J. Histogram matching augmentation for domain adaptation with application to multi-centre, multi-vendor and multi-disease cardiac image segmentation2020 December 01, 2020:[arXiv:2012.13871 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv201213871M.

[78] Vesal S, Gu M, Kosti R, Maier A, Ravikumar N. Adapt everywhere: unsupervised adaptation of point-clouds and entropy minimization for multi-modal cardiac image segmentation. IEEE Trans Med Imaging 2021;40(7):1838–51.

[79] Scannell C.M., Chiribiri A., Veta M. Domain-adversarial learning for multi-centre, multi-vendor, and multi-disease cardiac MR Image segmentation2020 August 01, 2020:[arXiv:2008.11776 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv200811776S.

[80] Chen C, Bai W, Davies RH, Bhuva AN, Manisty CH, Augusto JB, et al. Improving the generalizability of convolutional neural network-based segmentation on CMR images. Front Cardiovasc Med 2020;7:105.

[81] Chen X., Xia Y., Ravikumar N., Frangi A.F. Joint segmentation and discontinuity-preserving deformable registration: application to cardiac cine-MR images2022 November 01, 2022:[arXiv:2211.13828 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv221113828C.

[82] Wang H, Li Q, Yuan Y, Zhang Z, Wang K, Zhang H. Inter-subject registration-based one-shot segmentation with alternating union network for cardiac MRI images. Med Image Anal 2022;79:102455.

[83] Ma W., Yao H., Lin Y., Guo J., Li X. Semi-supervised domain generalization for cardiac magnetic resonance image segmentation with high quality pseudo labels 2022 September 01, 2022:[arXiv:2209.15451 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220915451M.

[84] Song L, Yi J, Peng J. LCC-Net: a lightweight cross-consistency network for semisupervised cardiac MR image segmentation. Comput Math Methods Med 2021;2021:9960199.

[85] Chen J., Li H., Zhang J., Menze B. Adversarial convolutional networks with weak domain-transfer for multi-sequence cardiac MR images segmentation2019 August 01, 2019:[arXiv:1908.09298 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv190809298C.

[86] Gu M., Vesal S., Kosti R., Maier A. Few-shot unsupervised domain adaptation for multi-modal cardiac image segmentation2022 January 01, 2022:[arXiv:2201.12386 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220112386G.

[87] Xiong Z, Xia Q, Hu Z, Huang N, Bian C, Zheng Y, et al. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Med Image Anal 2021;67:101832.

[88] Uslu F., Bharath A.A. TMS-Net: a segmentation network coupled with a run-time quality control method for robust cardiac image segmentation2022 December 01, 2022:[arXiv:2212.10877 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv221210877U.

[89] Hasan SMK, Linte CA. A multi-task cross-task learning architecture for Ad Hoc uncertainty estimation in 3D cardiac MRI image segmentation. Comput Cardiol (2010) 2021:48.

[90] Wang Y, Li ST, Huang J, Lai QQ, Guo YF, Huang YH, et al. Cardiac MRI segmentation of the atria based on UU-NET. Front Cardiovasc Med 2022;9:1011916.

[91] Yang G, Chen J, Gao Z, Li S, Ni H, Angelini E, et al. Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. Future Gener Comput Syst 2020;107:215–28.

[92] Ruijsink B, Puyol-Antón E, Li Y, Bai W, Kerfoot E, Razavi R, et al. Quality-aware semi-supervised learning for CMR segmentation. Stat Atlases Comput Models Heart 2020;2020:97–107.

[93] Chen W, Huang H, Huang J, Wang K, Qin H, Wong KKL. Deep learning-based medical image segmentation of the aorta using XR-MSF-U-Net. Comput Methods Progr Biomed 2022;225:107073.

[94] Li Y, Liu Z, Lai Q, Li S, Guo Y, Wang Y, et al. ESA-UNet for assisted diagnosis of cardiac magnetic resonance image based on the semantic segmentation of the heart. Front Cardiovasc Med 2022;9:1012450.

[95] Cheung W.K., Bell R., Nair A., Menezies L., Patel R., Wan S., et al. A computationally efficient approach to segmentation of the aorta and coronary arteries using deep learning. medRxiv. 2021 :2021.02.18.21252005.

[96] Li Z., Petri C., Howard J., Cole G., Varela M.PAT-CNN: Automatic segmentation and quantification of pericardial adipose tissue from T2-weighted cardiac magnetic resonance images2022 November 01, 2022:[arXiv:2211.04995 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv221104995L.

[97] Bui V, Hsu LY, Chang LC, Sun AY, Tran L, Shanbhag SM, et al. DeepHeartCT: a fully automatic artificial intelligence hybrid framework based on convolutional neural network and multi-atlas segmentation for multi-structure cardiac computed tomography angiography image segmentation. Front Artif Intell 2022;5:1059007.

[98] Zhao C., Shi S., He Z., Wang C., Zhao Z., Li X., et al. Spatial-temporal V-Net for automatic segmentation and quantification of right ventricles in gated myocardial perfusion SPECT images2021 October 01, 2021:[arXiv:2110.05443 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv211005443Z.

[99] Zhao Z., Hu J., Zeng Z., Yang X., Qian P., Veeravalli B., et al. MMGL: multi-scale multi-view global-local contrastive learning for semi-supervised cardiac image segmentation2022 July 01, 2022:[arXiv:2207.01883 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220701883Z.

[100] Huang Z, Gan Y, Lye T, Zhang H, Laine A, Angelini ED, et al. Heterogeneity measurement of cardiac tissues leveraging uncertainty information from image segmentation. Med Image Comput Comput Assist Interv 2020;12261:782–91.

[101] Abdulkareem M, Brahier MS, Zou F, Taylor A, Thomaides A, Bergquist PJ, et al. Generalizable framework for atrial volume estimation for cardiac CT images using deep learning with quality control assessment. Front Cardiovasc Med 2022;9:822269.

[102] Huang Z., Gan Y., Lye T., Liu Y., Zhang H., Laine A., et al. Cardiac adipose tissue segmentation via image-level annotations 2022 June 01, 2022:[arXiv:2206.04238 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220604238H.

[103] Li F, Sun L, Lam KY, Zhang S, Sun Z, Peng B, et al. Segmentation of human aorta using 3D nnU-Net-oriented deep learning. Rev Sci Instrum 2022;93(11):114103.

[104] Yoshida A, Kondo Y, Yoshimura N, Kuramoto T, Hasegawa A, Kanazawa T. U-Net-based image segmentation of the whole heart and four chambers on pediatric X-ray computed tomography. Radiol Phys Technol 2022;15(2):156–69.

[105] Park S, Chung M. Cardiac segmentation on CT Images through shape-aware contour attentions. Comput Biol Med 2022;147:105782.

[106] Jafari MH, Girgis H, Van Woudenberg N, Liao Z, Rohling R, Gin K, et al. Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training. Int J Comput Assist Radiol Surg 2019;14(6):1027–37.

[107] Zhu X, Wei Y, Lu Y, Zhao M, Yang K, Wu S, et al. Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging. Comput Methods Programs Biomed 2021;199:105914.

[108] Zhu X, Cheng Z, Wang S, Chen X, Lu G. Coronary angiography image segmentation based on PSPNet. Comput Methods Programs Biomed 2021;200:105897.

[109] Huang X., Chen Z., Yang X., Liu Z., Zou Y., Luo M., et al. Style-invariant cardiac image segmentation with test-time augmentation2020 September 01, 2020:[arXiv:2009.12193 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2020arXiv200912193H.

[110] Chartsias A., Joyce T., Papanastasiou G., Williams M., Newby D., Dharmakumar R., et al. Disentangled representation learning in cardiac image analysis2019 March 01, 2019:[arXiv:1903.09467 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv190309467C.

[111] Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans Med Imaging 2018;37(11):2514–25. https://doi.org/10.1109/tmi.2018.2837502. Nov.

[112] Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. Lect Notes Comput Sci 2018:120–9. https://doi.org/10.1007/978-3-319-75541-0_13. Mar 15.

[113] Zotti C, Luo Z, Lalande A, Jodoin PM. Convolutional neural network with shape prior applied to cardiac MRI segmentation. IEEE J Biomed Health Inform 2019;23(3):1119–28. https://doi.org/10.1109/jbhi.2018.2865450.

[114] Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin PM. Cardiac segmentation with strong anatomical guarantees. IEEE Trans Med Imaging 2019;39(11):3703–13. https://doi.org/10.1109/tmi.2020.3003240.

[115] Campello VM, Gkontra P, Izquierdo C, Martin-Isla C, Sojoudi A, Full PM, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&MS challenge. IEEE Trans Med Imaging 2021;40(12):3543–54. https://doi.org/10.1109/tmi.2021.3090082. Dec.

[116] Chen X, Lian C, Wang L, Deng H, Kuang T, Fung SH, et al. Diverse data augmentation for learning image segmentation with cross-modality annotations. Med Image Anal 2021;71:102060. Apr.

[117] Bui V, Hsu LY, Shanbhag SM, Tran L, Bandettini WP, Chang LC, et al. Improving multi-atlas cardiac structure segmentation of computed tomography angiography: a performance evaluation based on a heterogeneous dataset. Comput Biol Med 2020;125:104019. Sep.

[118] Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes 2022;15(1).

[119] Hatamizadeh A., Nath V., Tang Y. Swin U.N.E.T.R: Swin transformers for semantic segmentation of brain tumors in MRI images [Internet]. 2022 Arxiv. [cited 2023Apr25]. Available from: https://arxiv.org/pdf/2201.01266.pdf.

[120] Grzeszczyk M.K., Płotka S., Sitek A. Multi-task Swin transformer for motion artifacts classification and cardiac magnetic resonance image segmentation. Statistical atlases and computational models of the heart regular and CMRxMotion challenge papers. 2023Jan;:409–17.

[121] Hooper S., Wornow M., Seah Y.H., Kellman P., Xue H., Sala F., et al. Cut out the annotator, keep the cutout: better segmentation with. [Internet]. OpenReview. 2020 [cited 2023Apr25]. Available from: https://openreview.net/forum?id=bjkX6Kzb5H.

[122] Chaitanya K., Erdil E., Karani N., Konukoglu E. Contrastive learning of global and local features for medical image segmentation with limited annotations [Internet]. 2020 arXiv.org. [cited 2023Apr25]. Available from: https://arxiv.org/abs/2006.10511.

[123] Roth HR, Yang D, Xu Z, Wang X, Xu D. Going to extremes: weakly supervised medical image segmentation. Mach Learn Knowl Extr 2021;3(2):507–24.

[124] Home [Internet]. MONAI; 2023 [cited Apr25]. Available from: https://monai.io/.