Attri-VAE: attribute-based, disentangled and interpretable representations of medical images with variational autoencoders

Irem Cetin*, Oscar Camara*, Miguel A. González Ballester*a,b

aBCN MedTech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
bICREA, Barcelona, Spain

Abstract

Deep learning (DL) methods where interpretability is intrinsically considered as part of the model are required to better understand the relationship of clinical and imaging-based attributes with DL outcomes, thus facilitating their use in reasoning medical decisions. Latent space representations built with variational autoencoders (VAE) do not ensure individual control of data attributes. Attribute-based methods enforcing attribute disentanglement have been proposed in the literature for classical computer vision tasks in benchmark data. In this paper, we propose a VAE approach, the Attri-VAE, that includes an attribute regularization term to associate clinical and medical imaging attributes with different regularized dimensions in the generated latent space, enabling a better disentangled interpretation of the attributes. Furthermore, the generated attention maps explained the attribute encoding in the regularized latent space dimensions. The Attri-VAE approach analyzed healthy and myocardial infarction patients with clinical, cardiac morphology, and radiomics attributes. The proposed model provided an excellent trade-off between reconstruction fidelity, disentanglement, and interpretability, outperforming state-of-the-art VAE approaches according to several quantitative metrics. The resulting latent space allowed the generation of realistic synthetic data in the trajectory between two distinct input samples or along a specific attribute dimension to better interpret changes between different cardiac conditions.

Keywords: deep learning, interpretability, attribute regularization, variational autoencoder, cardiac image analysis

1. Introduction

Deep learning (DL) methods have recently shown great success in many fields, from computer vision (Pitale et al., 2021; Zhu et al., 2017; Goodfellow et al., 2014) to natural language processing (Wu et al., 2019; Deng and Liu, 2018), among numerous others. In addition, DL methods have started to dominate the medical imaging field (Shen et al., 2017), being used in a variety of medical imaging problems, such as segmentation of anatomical structures in the images (Bernard et al., 2018; Ronneberger et al., 2015; López-Linares et al., 2018), disease prediction (Io et al., 2019), medical image reconstruction (Higaki et al., 2020; Koller et al., 2020) and clinical decision support (Sanchez-Martinez et al., 2022). Despite achieving exceptional results, DL methods face challenges when applied to medical data regarding explainability, interpretability, and reliability because of their underlying black-box nature (Singh et al., 2020; McCrindle et al., 2021). Hence, the need for tools that investigate the interpretability in DL is also emerging in healthcare.

Recent reviews of interpretable DL can be found in (Singh et al., 2020; Barredo Arrieta et al., 2020; Molnár, 2022; Masis, 2021). Some methods have been proposed that employ backpropagation-based attention maps to either generate class activation maps that visualize the
regions with high activations in specific units of the network (Selvaraju et al., 2017) or saliency maps using gradients of the inputs with respect to the outputs (Simonyan et al., 2014; Kapishnikov et al., 2019). Other methods also proposed creating proxy models that focus on complexity reduction such as LIME (Ribeiro et al., 2016) or by approximating a value based on game theory optimal Shapley values to explain the individual predictions of a model (Lundberg and Lee, 2017). However, it is key to design models that are inherently interpretable, rather than creating post-hoc models to explain the black-box ones (Rudin, 2019).

Recently, models based on latent representations, such as variational autoencoders (VAE), have become powerful tools in this direction (Liu et al., 2020; Biffi et al., 2020), as their latent space is able to encode important hidden variables of the input data (Kingma and Welling, 2014). Especially, when dealing with data that contains different interpretable features (data attributes), it is interesting to see how and if these attributes have been encoded in the latent space. Even though the proposed approaches provide promising results, they have some limitations, one of which is that the encoded variables cannot be easily controlled; they mostly show an entangled behavior, meaning each latent factor maps to more than one aspect in the generative process (Bengio et al., 2013).

In order to bypass this limitation, much effort has been done to enforce disentanglement in the latent space (Higgins et al., 2017a; Kim and Mnih, 2018; Rubenstein et al., 2018; Chen et al., 2018a; Chartsias et al., 2019), being the majority of them unsupervised techniques (Bengio et al., 2013; Locatello et al., 2019). While many of these methods show good disentanglement performance, they are not only sensitive to inductive biases (e.g., choice of network, hyperparameters, or random seeds), but also some amount of supervision is necessary for learning effective disentanglement (Locatello et al., 2019). Moreover, since these methods are able to learn a factorized latent representation without attribute specification, they require a post-hoc analysis to determine how different attributes are encoded to different dimensions of the latent space (Pati and Lerch, 2021).

On the other hand, attribute-based methods aim to establish a correspondence between data attributes of interest and the latent space (Hadjeres et al., 2017; Lample et al., 2017; Bouchacourt et al., 2018; Pati and Lerch, 2021). However, these methods also have their drawbacks: some of them are limited to work only on certain types of data attributes (Lample et al., 2017); some impose additional constraints (Bouchacourt et al., 2018); very few of them are designed to work with continuous variables (Hadjeres et al., 2017; Pati and Lerch, 2021); some require differentiable computation of the attributes; and they are extremely sensitive to the hyperparameters (Hadjeres et al., 2017). However, (Pati and Lerch, 2021) have recently shown promising results for interpretability with their approach, associating each data attribute to a different regularized dimension of the latent space, which they have applied in the MNIST database for digit number recognition. The same approach was also employed as a post-processing step to generate interpretable and temporally consistent segmentations of echocardiography images (Painchaud et al., 2021).

In this paper, we propose an attribute-interpreter VAE (Attri-VAE), an approach based on attribute-based regularization (Pati and Lerch, 2021) in the latent space, for an enhanced interpretation of clinical and imaging attributes obtained from multi-modal sources. Additionally, the proposed approach also enables classification, e.g., to identify healthy vs. pathological cases. Furthermore, we incorporate gradient-based attention map computation (Liu et al., 2020) to generate explanations of the attributes that are encoded in the regularized latent space dimensions. The main contributions of this work can be described as follows:

- The proposed approach is able to interpret different data attributes where specific ones are forced to be encoded along specific latent dimensions without the need for any post-hoc analysis, while encouraging attribute disentanglement by employing β-VAE as a backbone (Higgins et al., 2017a).

- The structured latent space enables controllable data generation by changing the latent code of the regularized dimension (i.e., following the corresponding attribute), generating new data samples as a result of manipulating these dimensions. For instance, if the attribute represents volume in a region of interest (ROI) and the corresponding regularized dimension is the first one of the latent code, then increasing values of the dimension would result in increasing the ROI volume.
• Attribute-based gradient-based attention maps provide a way to explain how the gradient information of individual attributes flow inside the proposed architecture.

• The classification network provides a way to stratify different cohorts, based on the attributes in the latent space. In this way, the most discriminative features for the classification task are identified by projecting original samples into the latent space.

In this work, we have applied the proposed Attri-VAE approach to study cardiovascular pathological conditions, such as myocardial infarction, using the EMIDEC cardiac imaging dataset (Lalande et al., 2020), including clinical and imaging features, also exploring the association with radiomics descriptors. Additionally, we used ACDC MICCAI17 database\(^1\) as an external testing dataset.

The remainder of this paper is organized as follows. Firstly, we present the methodology and the details of our architecture in Section 2. We then describe the experimental setup and employed dataset in Section 3. Section 4 provides the results that are discussed in Section 5. Finally, in Section 6 we conclude our findings.

2. Methodology

The overall structure of our framework is shown in Figure 1 (training) and Figure 2 (testing). The proposed Attri-VAE incorporates attribute regularization into a \(\beta\)-VAE framework that was used as a backbone for the interpretation of data attributes. The trained network enables to generate new data samples by manipulating the data attributes, whereas the generated attribute-based attention maps explain how the gradient information of each attribute flows inside the proposed architecture. This section is organized firstly explaining the overall training criterion of the proposed model, with the following subsections describing each of the elements of our methodology and their integration.

2.1. Training criterion

Attri-VAE is trained with a loss function, \(\mathcal{L}\), which is composed of four terms, as follows:

\[
\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{MLP}} + \gamma \mathcal{L}_{\text{AR}}. \tag{1}
\]

The reconstruction loss, \(\mathcal{L}_{\text{recon}}\), is based on the binary cross-entropy (BCE) between the input \(X\) and its reconstruction \(\hat{X}\), while the second term, \(\mathcal{L}_{\text{KL}}\), employs the Kullback-Leibler (KL) divergence between the learned prior and the posterior distributions, weighted by a hyperparameter (\(\beta\)). An additional term, \(\mathcal{L}_{\text{MLP}}\), estimates the BCE loss for the classification between the network prediction, \(y_{\text{GT}}\), and the ground truth label, \(y_{\text{GT}}\). The final loss term, \(\mathcal{L}_{\text{AR}}\), includes the attribute regularization, with a tunable hyperparameter (\(\gamma\)) that weights its strength. In the following sections, detailed explanations of each loss term in our training criterion can be found (also see Figure 1).

2.2. Variational autoencoder (VAE) and \(\beta\)-VAE

A variational autoencoder (Kingma and Welling, 2014) is a generative model that consists of an encoder and a decoder. The encoder, \(q_\phi(Z|X)\), approximates the posterior distribution with parameters \(\varphi\), taking as input \(X\) from a high dimensional space, and learning to map it onto a low dimensional space by outputting the mean and variance (\(\mu\) and \(\sigma\), respectively) of a Gaussian probability density. The resulting low dimensional space is referred to as a latent space, with points \(Z\) in the latent space being the latent vectors. The decoder, \(p_\theta(X|Z)\), parameterized by \(\theta\), takes a latent vector \(Z\) that is sampled from \(p(Z)\) (prior distribution, e.g., unit Gaussian), using the reparameterization trick (Kingma and Welling, 2014), and outputs \(\hat{X}\), which is a reconstructed version of the input \(X\).

A variational autoencoder aims to maximize the marginal likelihood of the reconstructed output, which is written as:

\[
\log p_\theta(X) \geq \mathbb{E}_{Z \sim q_\phi(Z|X)}[\log p_\theta(X|Z)] - D_{KL}(q_\phi(Z|X)||p(Z)) \tag{2}
\]

In this objective function, the first term is the log likelihood expectation that the input \(X\) can be generated by the
Figure 1: Training framework of the proposed approach. Loss functions are shown in red arrows. The total loss function of the model is:
\[ \mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{MLP}} + \gamma \mathcal{L}_{\text{AR}} \]

(a) Losses computed for each data sample: multilayer perceptron (MLP) loss (\(\mathcal{L}_{\text{MLP}}\)), Kullback-Leibler (KL) loss (\(\mathcal{L}_{\text{KL}}\)), and reconstruction loss (\(\mathcal{L}_{\text{recon}}\)). (b) Attribute-regularization loss (\(\mathcal{L}_{\text{AR}}\)), computed inside a training batch that has \(n\) data samples.

The input, a 3D image (\(X\)), first goes through the 3D convolutional encoder, \(q_\phi(Z|X)\), which learns to map \(X\) to the low dimensional space \(Z\) by outputting the mean (\(\mu\)) and variance (\(\sigma\)) of the latent space distributions. The decoder, \(p_\theta(\hat{X}|Z)\), then takes \(Z\) and outputs the reconstruction of the original input, (\(\hat{X}\)). The predicted classes of the inputs, \(y_c\), are computed with a MLP module that consists of three fully connected (FC) layers. The corresponding MLP loss function is computed between \(y_c\) and the ground truth label \(y_{GT}\). In (b), \(\mathcal{L}_{\text{AR}}\) is shown to regularize the first dimension of the latent space (\(Z_1\)) with the attribute \(a_1\). \(\mathcal{L}_{\text{AR}}\) is computed inside a training batch that has \(n\) data samples.

sampled \(Z\) from the inferred distribution, \(q_\phi(Z|X)\). The second term corresponds to the KL divergence between the distribution of \(Z\) inferred from \(X\), and the prior distribution of \(Z\). Note that both distributions are assumed to follow a multivariate normal distribution.

In practice, the loss function of the VAE consists of two terms: a first term that penalizes the reconstruction error between the input and output; and a second term...
For a given 3D data sample, $X$, the trained 3D convolutional encoder, $q_\phi(Z|X)$, outputs the mean ($\mu$) and variance ($\sigma$) vectors, then $Z$ being sampled with the reparameterization trick. (a) Data generation process by changing only first ($Z_1$) and second ($Z_2$) regularized latent dimensions of $Z$, which correspond to two different data attributes (volume and maximum 2D diameter, respectively). Then, the decoder, $p_\theta(X|Z)$, generates 3D outputs, $X_1$ and $X_2$, using the manipulated latent vectors, $Z_1$ and $Z_2$, respectively. (b) Attribute-based attention map generation for a given attribute, which is encoded in the first latent dimension ($Z_1$). First, ($Z_1$) is backpropagated to the encoder’s last convolutional layer to obtain the gradient maps ($Grads_1$ and $Grads_2$) with respect to the feature maps ($F_1$ and $F_2$). The gradient maps of ($Z_1$) measure the linear effect of each pixel in the corresponding feature map on the latent values. After that, we compute the weights ($w_1$ and $w_2$) with global average pooling (GAP) on each gradient map. A heat map is generated by multiplying these values ($w_1$ and $w_2$) with the corresponding feature map and summing them up and applying an activation unit (ReLU). Additionally, the class score of the input, $y_c$, is computed with the multilayer perceptron (MLP) that is connected to $Z$. Note that, in the figure it is assumed that the last convolutional layer of the encoder has 2 feature maps.

forcing the learned distribution, $q_\phi(Z|X)$, to be as similar as possible to the prior distribution, $p(Z)$. In this case, the overall VAE loss can be written as:

$$\mathcal{L}_{VAE}(\theta, \varphi) = \mathcal{L}_{recon}(\theta, \varphi) + \mathcal{L}_{KL}(\theta, \varphi),$$

where the reconstruction loss, $\mathcal{L}_{recon}(\theta, \varphi)$, and the KL loss, $\mathcal{L}_{KL}(\theta, \varphi)$, are computed as follows:
\[ L_{\text{recon}}(\theta, \varphi) = \sum_{i=1}^{N} \| \hat{X} - X \|^2, \]  
\[ L_{\text{KL}}(\theta, \varphi) = D_{KL}(q_\varphi(Z|X)||p(Z)). \]

When \( q_\varphi(Z|X) \) is a multivariate normal distribution with parameters \( \mu \) and \( \sigma^2 \), the objective loss function is differentiable with respect to \((\theta, \varphi, \sigma, \mu)\) (Kingma and Welling, 2014), and the parameters of the VAE can be optimized iteratively with stochastic gradient descent algorithms (Kingma and Ba, 2015).

A latent representation is disentangled if each dimension in the latent space is sensitive to one generative factor and comparably invariant to the changes in the other factors (Liu et al., 2021). Such a disentangled representation is a great asset for interpretability. In this work we chose to use \( \beta \)-VAE as the backbone of our approach to encourage the disentanglement as it is easy to formulate and it has shown good performance based on one or more disentanglement metrics (Higgins et al., 2017a; Burgess et al., 2018).

The \( \beta \)-VAE approach (Higgins et al., 2017a) is an extension of the standard VAE that aims to learn a disentangled representation of the encoded variables in a completely unsupervised manner (Locatello et al., 2019; Higgins et al., 2017a) by simply giving more weight to the KL term, compared to the original VAE, with an extra hyperparameter \( \beta \):

\[ L_{\text{VAE}}(\theta, \varphi) = L_{\text{recon}}(\theta, \varphi) + \beta L_{\text{KL}}(\theta, \varphi), \]

The main idea here is that adding \( \beta \) restrains the latent representation, forcing it to be more factorized (Higgins et al., 2017a; Burgess et al., 2018); when \( \beta > 1 \), it encourages dimensional independence in the latent space, hence leading to a better disentanglement. On the other hand, when \( \beta = 1 \), it becomes equivalent to the standard VAE. Although, higher values of \( \beta \) have shown promising results to encourage disentangling (Higgins et al., 2017b), they often lead to a trade-off between reconstruction accuracy and the disentanglement of the latent space. For this reason, a well chosen \( \beta \) is necessary for both reconstruction accuracy and disentanglement.

### 2.3. Attribute-based regularization

In order to better interpret the data attributes that are encoded in the latent space, we employ an attribute-based regularization loss (Pati and Lerch, 2021), which aims to encode an attribute \( a \) along a dimension \( d \) of the latent space (regularized dimension). In this way, as one interpolates along dimension \( d \) (in a \( D \)-dimensional latent space), the attribute value of the generated data is also monotonically changed. Therefore, our hypothesis is that a model trained with an attribute-based regularization not only improves interpretation but also can be used to generate controllable images by manipulating different dimensions of the latent space, which are corresponding to different data attributes.

In this sense, the attribute regularization loss, \( L_{\text{AR}} \), is calculated for the dimension \( d \) of the latent space in a training batch containing \( n \) training examples for the purpose of forcing the dimension \( d \) to have a monotonic relationship with the attribute values of \( a \). The attribute regularization loss is then computed as follows:

\[ L_{\text{AR}}(d, a) = \text{MAE}(\tanh(\delta \text{Dist}_Z) - \text{sgn} (\text{Dist}_a)), \]

where \( \text{MAE} \) is the mean absolute error, \( \text{Dist}_a \) is the attribute distance matrix, and \( \text{Dist}_Z \) is the distance matrix of the latent dimension \( d \). These matrices are computed for all \( n \) data examples in the corresponding training batch, such that:

\[ \text{Dist}_a = a(X_i) - a(X_j), \]
\[ \text{Dist}_Z = Z_{ij} - Z_{ij}, \]

where \( i, j \in \{0, n\}, X_i \) and \( X_j \) are two exemplary samples (Equation 8), and each \( D \)-dimensional latent vector is represented as \( Z = [Z^d] \), where \( d \in \{0, D\} \) (Equation 9).

In Equation 7, \( \tanh \) and \( \text{sgn} \) refer to hyperbolic tangent function and sign function, respectively, whereas \( \delta \) is the hyperparameter that modulates the spread of the posterior distribution. As we are interested in whether a certain sample’s attribute value is higher or lower than the others inside the corresponding mini-batch, the \( \text{sgn} \) function is used. Additionally, a \( \tanh \) function was chosen for the regularized dimension’s distance matrix, \( \text{Dist}_a \), because it has the same range as \( \text{sgn}(\text{Dist}_a) \), and it is a differentiable function (i.e., the loss is also differentiable with respect to...
the latent vectors and the encoder’s parameters). Consequently, the objective function tries to minimize the MAE between \( \text{tanh}(\delta \text{Dist}_p) \) and \( \text{sgn}(\text{Dist}_a) \) so that the regularized dimension has a monotonic relationship with the attribute values.

While the above procedure gives an objective function for one attribute, for multiple selected attributes of interest to be encoded in the latent space, the overall loss function can be computed by summing all the corresponding objective functions together. Specifically, when the attribute set is \( A : \{a_k\} \), where \( k \in [0, K) \) contains \( K \) attributes (\( K \leq D \), being \( D \) the latent size), then the overall loss function is computed as:

\[
L_{AR} = \sum_{k=0}^{K-1} L_{d_k,a_k},
\]

where \( d_k \) represents the index of the regularized dimension for the attribute \( k \). This process is represented in Figure 1(b).

### 2.4. Classification network

Recently, performing a classification task using VAEs has been proposed to learn and separate different cohorts in the latent space. For example, Bifﬁ et al. [Bifﬁ et al., 2020] classified heart pathologies with cardiac remodelling using explainable task-speciﬁc shape descriptors learned directly with a VAE architecture from the input segmentations. Additionally, other approaches based on VAE have also been applied to analyse coronary artery diseases [Clough et al., 2019], Alzheimer’s disease [Shakeri et al., 2016] or to predict the response of cardiomyopathy patients to cardiac resynchronization therapy [Puyol-Antón et al., 2020].

In this line, to enforce class separation to the Attri-VAE, a multilayer perceptron (MLP) prediction network was connected to the latent vector, \( p(y_c | Z) \) (see Figure 1). The corresponding objective function can be computed as the binary cross entropy (BCE) between the network prediction \( y_c \) and the ground truth label \( y_{GT} \), such that:

\[
L_{MLP} = \text{BCE}(y_c, y_{GT})
\]

### 2.5. Attribute-based attention generation

The Attri-VAE facilitates data interpretation by generating new data samples as a result of scanning the regularized latent dimensions. Furthermore, it also provides a way to obtain attention maps from these dimensions (attribute-based attention map generation) for a better understanding on how gradient information of these attributes flows inside the proposed architecture (as can be seen in Figure 2).

Attribute-based visual attention maps were generated by means of gradient-based computation (Grad-CAM) [Selvaraju et al., 2017], as proposed by (Liu et al., 2020). Basically, a score is calculated from the latent space that is then used to estimate the gradients and attention maps. Specifically, given the posterior distribution inferred by the trained network for a data sample \( X_q, q(Z|X) \), the corresponding \( D \)-dimensional latent vector \( Z \) is sampled using the reparameterization trick [Kingma and Welling, 2014]. Subsequently, for a given attribute set \( A : \{a_k\} \), where \( k \in [0, K) \) contains \( K \) attributes, attribute-based attention maps, \( M_{d_k} \), are generated for each regularized latent dimension \( Z_{d_k} \) by backpropagating the gradients to the encoder’s last convolutional feature maps \( F : \{F_i\} \) where \( i \in [0, n) \):

\[
M_{d_k} = \text{ReLU} \left( \sum_{i=1}^n w_i F_i \right),
\]

where \( d_k \) is index of the regularized latent dimension for a given attribute \( k \). The weights, \( w_i \), are computed using global average pooling (GAP), which allows us to obtain a scalar value, as follows:

\[
w_i = \text{GAP} \left( \frac{\partial Z_{d_k}}{\partial F_i} \right) = \frac{1}{T} \sum_{p=1}^j \sum_{q=1}^l \frac{\partial Z_{d_k}}{\partial F_{pq}},
\]

where \( T = j \times l \), (\( i.e., width \times height \)), and \( F_{pq} \) is the pixel value at location \((p, q)\) of the \( j \times l \) matrix \( F_i \). This process is visually summarized in Figure 2.

### 3. Application for interpretable cardiology

#### 3.1. Datasets

Initially, the EMIDEC dataset [Lalande et al. 2020] was used in our experiments. It is a publicly available database with of delay-enhancement magnetic resonance images (DE-MRI) of 150 cases (100 and 50 cases for training and testing, respectively), with the corresponding
clinical information. Each case includes a DE-MRI acquisition of the left ventricle (LV), covering from base to apex. The training set, with ground-truth segmentations, includes 67 myocardial infarction (MINF) cases and 33 healthy subjects. The testing set includes 33 MINF and 17 healthy subjects. Some clinical parameters were also provided along with the MRI: sex, age, tobacco (yes, no, and former), overweight, arterial hypertension, diabetes, family history of coronary artery disease, electrocardiogram (ECG), Killip max

3

4

5

6

Furthermore, we also used an additional external testing dataset for a more robust assessment of the classification performance, the ACDC MICCAI17 challenge training dataset

6

(https://www.creatis.insa-lyon.fr/Challenge/acdc/). The ACDC dataset includes ground-truth segmentations of the left ventricle, myocardium and right ventricle by an experienced manual observer at both end-diastole (ED) and end-systole (ES) cine-MRI from 20 healthy volunteers and 20 MINF cases). The ACDC dataset includes ground-truth segmentations of the left ventricle, myocardium and right ventricle by an experienced manual observer at both ED and ES timepoints (Bernard et al., 2018). The reader is referred to (Lalande et al., 2020; Bernard et al., 2018) for more details on the MRI acquisition protocol.

As a pre-processing step, the intensities of the left ventricle in all images were scaled between 0 and 1. Additionally, each image was cropped and padded ($x = 80; y = 80; z = 80; t = 1$).

3.2. Cardiac attributes

Three different types of attributes were studied in our experiments. Initially, the Attri-VAE was trained with cardiac shape descriptors (e.g., wall thickness, LV and myocardial volumes, ejection fraction), extracted from ground-truth segmentations, which can easily be visually interpreted. In addition, attributes available from clinical information with the highest discriminative performance were identified using recursive feature elimination (RFE) with a support vector machines (SVM) classification model (linear kernel, regularization parameter $C = 10$) since this approach has already shown good performance for feature selection tasks (Huang et al., 2014; Samb et al., 2012; Yang, 2018). The most discriminative attributes were then included in our analysis (e.g., gender, age, tobacco). The feature selection pipeline was done using the python-based machine learning library scikit-learn (version 1.0.2).

Finally, the Attri-VAE was also trained with radiomics features. Radiomics analysis was originally proposed to capture alterations at both the morphological and tissue levels in oncology applications (Aerts et al., 2014; Lambin et al., 2017), deriving multiple quantifiable features from pixel-level data. More recently, radiomics approaches have provided promising results on cardiac MRI data, for discriminating different cardiac conditions (Neisius et al., 2019; Larroza et al., 2018; Baessler et al., 2018; Cetin et al., 2017), and to study cardiovascular risk factors in large databases (Cetin et al., 2020). Radiomics analysis represents a step towards interpretability compared to other black-box approaches since some features can be related to pathophysiological mechanisms (Cetin et al., 2020). However, there is a need for improving robustness and reproducibility of radiomics outcomes across different feature selection strategies and imaging protocols, which would lead to enhanced explainability. For this reason, radiomics features were employed in our experiments to benefit from the proposed network’s ability to explain the encoded attributes. The open source library PyRadiomics (version 3.0.1) was used to derive 114 features per analysed cardiac structure. Subsequently, radiomics features with the highest discriminative performance were identified using the above-mentioned feature selection approach as this strategy has also demonstrated good performance with previous radiomics studies (Zhang et al., 2017; Xiao et al., 2020; Chen et al., 2018b). The top performing features of this process were then selected to train the Attri-VAE.

3.3. Architectural details

The 3D convolutional encoder of the proposed Attri-VAE framework compresses the input into a 250 dimensional embedding through a series of 3D convolutional

\cite{scikit-learn

\cite{pyradiomics.readthedocs.io/}
layers with kernel size 3 and stride 2, except the last convolutional layer that has stride 1. The prediction network was constructed with a shallow 3-layer MLP to be able to discriminate between the healthy and infarct subjects, using a ReLU activation function as a non-linearity after the first two layers. The upsampling and convolutional layers used in the encoder and the decoder were followed by batch normalization and ReLU non-linearity, except the decoder’s last convolutional layer (Attri-VAE output) where a sigmoid function was applied. All the network weights were randomly initialized with Xavier initialization (Glorot and Bengio, 2010). The tunable parameters of the loss function (Equation 1) were fixed as follows: KL weight $\beta = 2$; and regularization weight $\gamma = 200$. Additionally, $\delta$ (Equation 2) was set to 10. The model architecture and other details are provided in our GitHub repository.

The Attri-VAE was trained on a NVIDIA Tesla T4 GPU using Adam optimizer with learning rate equals to 0.0001 and batch size of 16 for 10000 epochs. The dataset was splitted into 70/30 training (47 pathological, 23 healthy) and testing (20 pathological and 10 healthy subjects) sets. Subsequently, random oversampling of the normal subjects was employed in the training set as a strategy to treat the unbalanced behavior of the dataset; however, testing set was kept unchanged. Note that the proposed model is implemented using python programming language and PyTorch library (version 1.10.0). Image pre-processing and transformations were done using the python-based MONAI library (version 0.8.0).

3.4. Experimental setting and evaluation criteria

The performance of the proposed Attri-VAE, both qualitatively and quantitatively, was compared with baseline VAE and $\beta$-VAE models in several experiments. First of all, the degree of disentanglement of the proposed latent space was evaluated with respect to different data attributes, using the following metrics available in the literature: the modularity metric, to analyse the dependence of each dimension of the latent space on only one attribute (Ridgeway and Mozer, 2018); the mutual information gap (MIG), to evaluate the MI difference between a given attribute and the top two dimensions of the latent space that share maximum MI with the corresponding attribute (Chen et al., 2018a); the separated attribute predictability (SAP), to measure the difference in the prediction error of the two most predictive dimensions of the latent space for a given attribute (Kumar et al., 2017); and the Spearman correlation coefficient (SCC) score, to compute its maximum value between an attribute and each dimension of the latent space.

In parallel, the interpretability metric introduced in (Adel et al., 2018) was used to measure the ability to predict a given attribute using only one dimension of the latent space. As for the $\beta$-VAE mode, dimensions having a high MI with the corresponding data attribute were chosen for the interpretability estimation. The reconstruction fidelity performance was also evaluated, employing the maximum mean discrepancy (MMD) score (Gretton et al., 2007), which measures the distance between the distributions of real and reconstructed data examples, as well as their mutual information (MI) as an image similarity metric. The interpretability and MI metrics were then used to identify the optimal values of the most relevant hyperparameters in Equation 10 and Equation 7 (i.e., $\beta$, $\gamma$ and $\delta$), evaluating the influence of the KL divergence ($\beta$) and attribute regularization ($\gamma$) loss terms, as well as the weight of the distance matrix between two samples in a latent dimension. As a proof-of-concept, the hyperparameter sensitivity analysis was performed with only the four cardiac shape-based interpretable attributes.

Another set of experiments was carried out to explore the potential of the latent space generated by the Attri-VAE approach to create synthetically realistic samples. First, two samples in the Attri-VAE latent space, corresponding to input data with distinct cardiac characteristics (e.g., thin vs thick myocardium, absence vs presence of myocardium infarct), were chosen as references to synthetically generate interpolated images through their trajectory. Secondly, we qualitatively evaluated the control over individual data attributes during the generation process of the Attri-VAE model. Given a sample with a latent code $z$, a given attribute (e.g., LV volume) can be scanned from low to high values changing the latent code of the corresponding regularized dimension, due to their monotonic relationship. The attribute scanning creates synthetically generated samples in a latent space trajectory where
only the chosen attribute is changing, facilitating its interpretation. In order to further facilitate the identification of each attribute’s visual influence in the synthetically generated images, gradient-based attention maps were also estimated.

Finally, the performance of the Attri-VAE model for classifying healthy and pathological hearts was assessed using the area under the curve (AUC) and accuracy (ACC) metrics, using both the EMIDEC and the ACDC17 challenge datasets. The Attri-VAE results were benchmarked against other VAE-type approaches (VAE+MLP, β-VAE+MLP), as well as to classical radiomics analysis (with SVM). The latent space projections of the Attri-VAE model, regularized by different attributes, were also qualitatively analysed to identify the attributes better differentiating healthy and pathological clusters of samples.

4. Results

4.1. Hyperparameter sensitivity analysis

Figure 3 shows the effect of several hyperparameters on the interpretability and reconstruction fidelity of the Attri-VAE scheme. For comparison, the performance of β-VAE (β = 3) is also represented. A visual inspection of the figure suggests that γ, i.e., the hyperparameter controlling the attribute regularization, was the key to obtain good interpretability values while keeping reasonable reconstruction fidelity (mutual information ≥ 0.88), with values of γ ≥ 100. Additionally, values of δ ≤ 10 (e.g., hyperparameter on the attribute regularization controlling the weight of the distance matrix between two samples) also ensured a good trade-off between interpretability and reconstruction fidelity. On the other hand, the β hyperparameter was not as relevant as the other two. As expected, the β-VAE approach without attribute regularization, provided acceptable reconstruction fidelity results but low values of interpretability. We need to point out that the same results were obtained when using radiomics features instead of shape-based attributes.

4.2. Disentanglement and interpretability

The proposed Attri-VAE approach outperformed β-VAE across all tested disentanglement metrics using shape and clinical attributes, implying a more disentangled latent space. Firstly, both Attri-VAE and β-VAE provided high modularity values (Attri-VAE: 0.98 vs. β-VAE: 0.97), signalling that each dimension of the latent spaces in both models only depended on one data attribute. The Attri-VAE also resulted in higher MIG/SAP scores than β-VAE (Attri-VAE: 0.49/0.01 vs. β-VAE: 0.05/0.06). In its turn, the SCC metric estimated for Attri-VAE was substantially higher than the corresponding β-VAE one (Attri-VAE: 0.97 vs. β-VAE: 0.46) due to the monotonic relationship between a given attribute and the regularized latent dimension enforced by the former. When using radiomics features, the same trend was observed, with some Attri-VAE disentanglement metrics (MIG and SAP) slightly lower than when using shape and clinical attributes (Attri-VAE / β-VAE): modularity, 0.98/0.98; MIG, 0.49/0.01; SAP, 0.51/0.06; SCC, 0.98/0.42).

Table 1 shows the interpretability scores for both Attri-VAE and β-VAE obtained with shape, clinical and radiomics attributes. The radiomics feature selection identified seven of them having the most discriminative power: four shape-based, being the sphericity of the left ventricle, the maximum 2D diameter of the myocardium, as well
Figure 4: Three examples of real and reconstructed images using the VAE, \( \beta \)-VAE and Attri-VAE approaches. Three slices are shown in every example: apical (APEX), mid-ventricle (MID) and basal (BASE) slices. Sample 1 and 3 correspond to healthy hearts while Sample 2 shows an infarcted myocardium.

Table 1: Interpretability score \([\text{Adel et al., 2018}]\) of most relevant shape, clinical and radiomics attributes, as encoded in the latent space, with the Attri-VAE and \( \beta \)-VAE approaches. LV: left ventricle, MYO: myocardium, EF: ejection fraction. Maximum interpretability is 1.0.

| Attribute     | Attri-VAE | \( \beta \)-VAE |
|---------------|-----------|----------------|
| LV volume     | 0.89      | 0.14           |
| MYO volume    | 0.93      | 0.02           |
| Wall thickness| 0.95      | 0.10           |
| EF            | 0.94      | 0.03           |
| Gender        | 0.98      | 0.19           |
| Age           | 0.93      | 0.12           |
| Tobacco       | 0.70      | 0.19           |
| Radiomics     | 0.91      | 0.06           |

as left ventricle and myocardial volumes; three texture-based, being the correlation of the left ventricle, the difference entropy of the myocardium and the inverse variance of the left ventricle.

It can easily be observed that the Attri-VAE provided a high degree of interpretability (i.e., close to 1.0) for all attributes, with the exception of tobacco (0.70). Among shape and clinical features, gender was the attribute with a higher interpretability (0.98), followed by the wall thickness (0.95), meaning that they could be predicted with only one dimension of the latent space. As for radiomics features, the average interpretability metric value was of 0.91, with shape-based ones showing slightly larger values than texture features (0.93 and 0.89, respectively); the maximum 2D diameter of the myocardium presented the highest value (0.97). On the other hand, the \( \beta \)-VAE clearly resulted in lower interpretability values (average of 0.11 for shape/clinical attributes and 0.06 for radiomics features).

### 4.3. Reconstruction fidelity

Table 2 summarizes the results of the reconstruction fidelity metrics (MMD and MI) for the VAE, \( \beta \)-VAE and Attri-VAE models. The proposed Attri-VAE approach obtained the lowest MMD values, representing a lower distance between input and reconstructed images. However, the VAE approach had the (slightly) best MI (0.91 and 0.89 for VAE and Attri-VAE, respectively), since its latent space was less constrained, compared to the other models.

Figure 5 shows the reconstructions of three data examples from the EMIDEC dataset using the VAE, \( \beta \)-VAE and Attri-VAE approaches. Even though the three models achieved similar qualitative reconstruction results, the Attri-VAE model generated images better preserving the
heart shape and details than the other models: see the papillary muscles in mid-myocardium slices (dark regions in the blood pool) or the left ventricular cavity in apical slices of Sample 2 and Sample 3 in Figure 4. We can also observe in the figure that apical slides were more difficult to reconstruct than mid-ventricle and basal ones for the three tested models.

4.4. Latent space interpolation and attribute scanning

Figure 6 shows three examples of interpolation between two distinct and well-separated samples in the learned latent space of the Attri-VAE model. As it can be appreciated in the figure, the proposed approach generates synthetic interpolated images that have a realistic appearance, gradually changing the main sample characteristics in the trajectory between the chosen samples. The first row of Figure 6 clearly demonstrate the Attri-VAE model’s ability to create smooth transitions between hearts having largely different characteristics such as (thin to thick) wall thickness. The other two rows of the figure demonstrate a similar behaviour from non-infarcted/scar to infarcted/scar patients.

Figure 5 illustrates the effect of scanning an individual attribute along its corresponding regularized dimension in the Attri-VAE model, where all the remaining attributes remain fixed. The first three rows of the figure exemplify the attribute scanning that was done on the latent space of Attri-VAE, which was trained with clinical plus shape features. The rest of the rows represent the attribute scanning on the latent space of Attri-VAE trained with selected radiomics features. For shape-based attributes, the changes in the attribute when moving along different values of the regularized dimension are clearly seen. For instance, from the left to the right in Figure 5 how LV and myocardial volumes are increasing in the first and second rows, respectively, or how the LV becomes more spherical. More subtle changes are observed with texture-based radiomics but they can still be identified with a careful inspection of the generated images. For example, moving along the latent space dimension corresponding to the correlation LV, we find more or less intensity homogeneity in the LV. The LV inverse variance (LV-IV) and the difference entropy of the myocardium (DE-MYO) only produced small changes that consisted in slightly thicker myocardium with lower values of LV-IV (left samples in Figure 5) and some more darker patches and heterogeneous texture in the myocardium for higher values of DE-MYO (right in Figure 5). It needs to be pointed out that attribute scanning for clinical attributes such as age, gender and tobacco is not shown since the images do not visually change along the corresponding regularized dimensions.

Additionally, the right side of Figure 5 shows the attention maps associated with the changes in each regularized dimension of the Attri-VAE model, as a way to better understand the effect of each studied attribute. We can see in the figure that more attention (i.e., higher response) is paid to more varying regions for shape-based attributes (e.g., right side of the slide for LV volume, where LV is increasing from the left to the right in the regularized dimension). In general, attention maps for texture-based features have less high-response regions than for shape-based attributes. However, in some texture-based features such as the difference entropy of the myocardium, higher response can still be localized (in this example, darker regions in the top left part of the slice). On the other hand, interpretation and validation of the resulting attention maps for other attributes such as for LV-IV are more challenging.

4.5. Classification

Table 3: Classification performance of EMIDEC and ACDC datasets (healthy vs. myocardial infarction) with different models. The results are reported as accuracy / AUC score. SVM: support vector machine.

Table: Reconstruction accuracy on the EMIDEC dataset of the VAE, β-VAE and Attri-VAE approaches, quantified with the maximum mean discrepancy (MMD) and mutual information (MI) metrics. The MMD results are given as ± standard deviation.

|                  | EMIDEC       | ACDC        |
|------------------|--------------|-------------|
| VAE              | 0.96 / 0.94  | 0.58 / 0.54 |
| β-VAE            | 0.98 / 0.96  | 0.59 / 0.52 |
| Attri-VAE (Clinical+Shape) | 0.77 / 0.75  | 0.60 / 0.61 |
| Attri-VAE (Radiomics) | 0.91 / 0.90  | 0.45 / 0.31 |
| β-VAE+MLP        | 0.87 / 0.80  | 0.54 / 0.35 |
| VAE+MLP          | 0.87 / 0.80  | 0.54 / 0.35 |
| Radiomics analysis (SVM) | 0.87 / 0.80  | 0.54 / 0.35 |
Figure 5: Scanning of attributes and corresponding gradient-based attention maps for shape and radiomics features. The image in the middle (4th column, in yellow frame) shows the original reconstructed image. DE: difference entropy, IV: inverse variance, Max 2D dia: maximum 2-dimensional diameter, LV: left-ventricle, MYO: myocardium. Note that the first three rows demonstrate the attribute scanning that was done on the latent space of Attri-VAE, which was trained with clinical and shape features. The remaining rows represent the attribute scanning on the latent space of Attri-VAE trained with selected radiomics features.

Figure 6: Linear latent space interpolation between two data samples (extremes of each row in yellow frames) from the EMIDEC dataset. Each row depicts the interpolation from the left to the right latent vector dimension. Top: from thin to thick myocardium. Middle: from a myocardium with scar to one without. Bottom: from healthy subject to a patient with a myocardial infarct.

Table 3 shows that the Attri-VAE approach, besides increasing interpretability, it also achieves a better classification performance comparing to state-of-the-art models. The best result was obtained in both EMIDEC and ACDC datasets with the Attri-VAE trained with radiomics features (accuracy of 0.98 and 0.59 for both datasets), while the standard radiomics+SVM analysis was the worst for EMIDEC (accuracy of 0.77) and the β-VAE+MLP for ACDC. There were only minor differences in the accuracy of the Attri-VAE method when trained with clinical and shape attributes or radiomics features. All the evaluated models, trained with the EMIDEC data, substantially dropped their performance when tested on the external ACDC dataset, specially the VAE-based approaches.

Finally, the latent space projections of different regularized latent dimensions are visualized in Figure 7 with plot axes representing the encoded data attributes. As it can be observed in the figure, our model is able to build several reduced dimensionality spaces, based on different attributes, where healthy and pathological cases (red and blue in the figure, respectively) can easily be
clustered. For instance, the maximum 2D diameter of the myocardium and the LV volume attributes correctly separate most samples into two clusters. Interestingly, despite Attri-VAE having poor control over clinical attributes such as age or gender, they also facilitate the construction of the latent spaces and sample discrimination, as can be seen in the gender-age plot of Figure 7.

Figure 7: Latent space projections of regularized dimensions for different clinical, shape and radiomics attributes. Each point in the graphs represent a healthy or a myocardial infarction patient (red and blue, respectively), LV: left-ventricle, MYO: myocardium, IV: inverse variance, DE: difference entropy, Max 2D dia: maximum 2-dimensional diameter.

5. Discussion

The analysis of medical data demands for interpretable methods. However, the majority of deep learning methods do not fulfill the minimum level of interpretability to be used in reasoning medical decisions (Sanchez-Martinez et al., 2022), being difficult to relate clinically and physiologically meaningful attributes with model parameters and outcomes. Fortunately, interpretable and explainable deep learning methods are starting to emerge. Models creating latent space representations, such as variational autoencoders, are promising but attributes are usually entangled in the resulting reduced dimensionality space, hampering its interpretation. In this work, we have presented the Attri-VAE approach that generates disentangled and interpretable representations where different types of attributes (e.g., clinical, shape, radiomics) are individually encoded into a given dimension of the resulting latent space.

The results obtained by the proposed Attri-VAE model based on disentanglement and interpretability metrics clearly outperformed the state-of-the-art β-VAE approach, indicating a high degree of disentanglement and a monotonic relationship between a given attribute and the corresponding regularized dimension. However, Attri-VAE values for some metrics such as the MIG and SAP, although substantially better than those of β-VAE, were far from the maximum (e.g., 1.0). The same trend was observed by (Pati and Lerch, 2021) in the MNIST (i.e., for digit number identification) dataset, suggesting that other latent dimensions, beyond the regularized ones, share a high MI with different attributes.

Hyperparameter selection was a key step to find the optimal Attri-VAE configuration providing an excellent trade-off between reconstruction fidelity, at the level of state-of-the-art alternatives, and interpretability; even though the Attri-VAE approach had a more constrained latent space, it generated reconstructions that are less smooth than other VAE models and more similar to the original input images. The most critical parameter to enforce interpretability was the weight of the attribute regularization loss term ($\gamma$ in Equation 1), together with the influence of the distance matrix between two samples in a latent dimension ($\delta$ in Equation 7). The Attri-VAE plot of reconstruction fidelity vs interpretability, shown in Figure 3 had the same pattern as the one obtained by (Pati and Lerch, 2021). Interestingly, their optimal $\gamma$ values were lower than ours ([5.0, 10.0] vs ≥ 100), likely due to the higher complexity of the cardiac MRI data and corresponding latent space compared to the MNIST dataset. On the other hand, the best $\delta$ values were the same in the two studies ([1.0, 10.0]).

One of the most interesting characteristics of the Attri-VAE approach is the ability of creating realistic synthetic data by sampling the created latent space and interpolating between different original reconstructed inputs, which can be very useful for controllable and attribute-based data augmentation of training datasets in machine learning applications. Scanning a regularized dimension of the latent space creates synthetic images where the corresponding attribute changes its values, as can easily be observed for shape descriptors (e.g., LV and myocardial volumes, wall thickness) in Figure 5. In addition, the
proposed approach allows a better understanding of some (texture-based) radiomics features, which are often difficult to interpret. However, clinical attributes such as age, gender or tobacco consumption, despite obtaining good interpretability scores, did not create visually different interpolated samples over the regularized dimensions. One potential reason is the difficulty of the attribute regularization to control binary attributes, as suggested by (Pati and Lerch [2021]). Furthermore, the studied clinical attributes cannot be disassociated from shape and image intensity variations (e.g., morphological changes of the heart with age), thus it is too restrictive to keep all attributes fixed except a clinical one. In consequence, more work is needed to better construct latent spaces where clinical information can be disentangled from other attributes.

The generated gradient-based attention maps contributed to locally identify the cardiac regions where the attributes were influencing, which was particularly useful for global attributes and for complex features such as the texture ones. However, we only employed the well-known Grad-CAM method, which could be complemented with additional interpretability methods (e.g., LIME and its variations [Ribeiro et al. [2016]]) to better understand the attribute effects on the latent space. Additionally, the reliability of attention maps still requires further investigation to assess its robustness and reliability with respect to data input and model parameter perturbations [Reyes et al. [2020]]. In parallel, enhanced 3D visualizations of the generated samples are needed to have an overall perspective of the cardiac differences, beyond 2D slice views of the resulting images.

The proposed Attri-V AE model also achieved excellent classification performance (healthy vs. myocardial infarction), outperforming the other VAEBased approaches, with slightly better results when trained with radiomics. When evaluated in the EMIDEC training dataset with ground-truth labels, the Attri-V AE approach provided accuracy results (0.98) equivalent to the best challenge participants reporting their performance on the same dataset (1.0 [Lourenço et al. [2021]], 0.95 [Shi et al. [2021]], 0.94 [Ivantsits et al. [2021]] and 0.90 [Sharma et al. [2021]]). For the testing EMIDEC dataset [Lalande et al. [2021]], the best participant method obtained a decreased accuracy (0.82, [Lourenço et al. [2021]], [Girum et al. [2021]]), increasing to 0.92 for the challenge organizers (Shi et al. [2021]). As for the ACDC dataset, which was tested as an external database (i.e., without considering it in training), classification accuracy was substantially reduced (0.59), being worst than results reported by challenge participants [Bernard et al. [2018]] (0.96) to classify between the different pathologies (not only between healthy and myocardial infarction). Therefore, further work is required to improve the generalization of the Attri-V AE model to unseen data, being more robust to different quality and imaging acquisition protocols, through domain adaptation techniques, using databases such as the M&Ms challenge [Campello et al. [2021]].

One limitation of the Attri-V AE approach, also acknowledged by Pati and Lerch [Pati and Lerch [2021]], is the dependence on the selection of the data attributes to train the model. An incorrect attribute selection could lead to undesired strong correlations of several attributes that will not ensure a monotonic relationship with the corresponding regularized dimension, leading to less attribute interpretation and reconstruction quality. However, the projection of original samples in latent spaces with regularized dimensions for different attributes (see Figure 7) could be used as an interpretable attribute selection, identifying the ones better separating the analyzed classes such as the maximum 2D diameter of the myocardium and the LV volume attributes in our experiments. Further work will focus on fully integrating advanced feature selection techniques with the Attri-V AE model, as well as exploring alternative interpretability methods (see the recent review of Salahuddin et al. [Salahuddin et al. [2022]]) to better understand the role of clinical and imaging attributes on medical decisions in cardiovascular applications.

6. Conclusions

We have presented a novel approach, referred to as Attri-V AE, which implements attribute-based regularization in a β-V AE scheme with a classification module for the purpose of attribute-specific interpretation, synthetic data generation and classification of cardiovascular images. The basis of the proposed Attri-V AE model is to structure its latent space for encoding individual data attributes to specific latent dimensions, being guided by an attribute regularization loss term. The resulting constrained latent space can be easily manipulated along its regularized dimensions for an enhanced interpretation of
different attributes. Additionally, the proposed approach improves the current state-of-the-art for classifying cardiovascular images and allows the visualization of the most discriminative attributes by projecting the trained latent space. Future work will be focused on improving the generalization of the trained Attri-VAE models to images with different acquisition characteristics.

7. Competing interests

The authors declare that they have no competing interests.

8. Author contributions

All authors participated in the analysis of the data, critical revision of the manuscript, and final approval of the submitted manuscript. IC, MAGB, OC contributed to study concepts, methods and underlying data collection. IC, MAGB, OC drafted the manuscript. IC, MAGB, OC designed the machine learning methods. IC performed the data pre-processing and data analysis.

9. Funding

This work was partly funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project).

10. Availability of data and materials

This research was conducted using the publicly available EMIDEC and ACDC datasets. These datasets can be accessed in http://emidec.com/dataset and https://www.creatis.insa-lyon.fr/Challenge/acdc/ We have also made our code publicly available and can be found in https://github.com/iremcetin/Attri-VAE

11. Ethical approval

The datasets employed in this study are publicly available sources. Thus, for the detailed information see http://emidec.com/ and https://www.creatis.insa-lyon.fr/Challenge/acdc/

References

Adel, T., Grahramani, Z., Weller, A., 2018. Discovering interpretable representations for both deep generative and discriminative models, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning. PMLR. pp. 50–59. URL: https://proceedings.mlr.press/v80/adel18a.html

Aerts, H.J., Velazquez, E.R., Leijenaar, R.T.H., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D.H.F., Hoebers, F.J., Rietbergen, M.M., Leemans, C.R., Dekker, A., Quackenbush, J., Gillies, R.J., Lambin, P., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Communications 5, 4006. doi:10.1038/ncomms5006

Baessler, B., Luecke, C., Lurz, J., Klingel, K., von Roeder, M., de Waha, S., Besler, C., Maintz, D., Gutberlet, M., Thiele, H., Lurz, P., 2018. Cardiac MRI Texture Analysis of T1 and T2 Maps in Patients with Infarctlike Acute Myocarditis. Radiology 289, 357–365. doi:10.1148/radiol.2019190101

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115. doi:10.1016/j.inffus.2019.12.012

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35, 1798—1828. doi:10.1109/tpami.2013.50

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadiri, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maiер-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Issum, I.
Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M., 2018. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Transactions on Medical Imaging 37, 2514–2525. doi:10.1109/TMI.2018.2837502.

Biffl, C., Cerrolaza, J.J., Tarroni, G., Bai, W., de Marvao, A., Oktay, O., Ledig, C., Fologoc, L.L., Kamnitas, K., Doumou, G., Duan, J., Prasad, S.K., Cook, S.A., O’Regan, D.P., Rueckert, D., 2020. Explainable anatomical shape analysis through deep hierarchical generative models. IEEE Transactions on Medical Imaging 39, 2088–2099. doi:10.1109/TMI.2020.2964499.

Bouchacourt, D., Tomioka, R., Nowozin, S., 2018. Multilevel variational autoencoder: Learning disentangled representations from grouped observations. Proceedings of the AAAI Conference on Artificial Intelligence 32. URL: https://ojs.aaai.org/index.php/AAAI/article/view/11867.

Burgess, C.P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., Lerchner, A., 2018. Understanding disentangling in $\beta$-VAE. CoRR abs/1804.03599. URL: http://arxiv.org/abs/1804.03599.

Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreño, M., Albiol, A., Kong, F., Shadden, S.C., Acero, J.C., Sundaresan, V., Saber, M., Elattar, M., Li, H., Menze, B., Khader, F., Haarburger, C., Scannell, C.M., Veta, M., Carscadden, A., Punithakumar, K., Liu, X., Tsafaris, S.A., Huang, X., Yang, X., Li, L., Zhuang, X., Viladés, D., Descalzo, M.L., Guala, A., Mura, L.L., Friedrich, M.G., Garg, R., Lebel, J., Henriques, F., Karakas, M., Çavuş, E., Petersen, S.E., Escalera, S., Segui, S., Rodríguez-Palomares, J.F., Lekadir, K., 2021. Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. IEEE Transactions on Medical Imaging 40, 3543–3554. doi:10.1109/TMI.2021.3090082.

Cetin, I., Raisi-Estabragh, Z., Petersen, S.E., Napel, S., Pieknik, S.K.P., Neubauer, S., Gonzalez Ballester, M.A., Camara, O., Lekadir, K., 2020. Radiomics signatures of cardiovascular risk factors in cardiac mri: Results from the uk biobank. Frontiers in Cardiovascular Medicine 7. doi:10.3389/fcvm.2020.591368.

Cetin, I., Sanroma, G., Petersen, S.E., Napel, S., Camara, O., Gonzalez Ballester, M.A., Lekadir, K., 2017. A radiomics approach to computer-aided diagnosis with cardiac cine-mri, in: Pop, M., Sermesant, M., Jodoin, P., Lalande, A., Zhuang, Y., Yang, G., Young, A.A., Bernard, O. (Eds.), Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges - 8th International Workshop. STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers. Springer. pp. 82–90. URL: https://doi.org/10.1007/978-3-319-75541-0_9.

Chartsias, A., Joyce, T., Papanastasiou, G., Semple, S., Williams, M., Newby, D.E., Dharmakumar, R., Tsafaris, S.A., 2019. Disentangled representation learning in cardiac image analysis. Medical Image Analysis 58, 101535.

Chen, R.T.Q., Li, X., Grosse, R.B., Duvenaud, D.K., 2018a. Isolating sources of disentanglement in variational autoencoders, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2018/file/1ee3dfcd8a0645a25a35977997223d22-Paper.pdf.

Chen, W., Liu, B., Peng, S., Sun, J., Qiao, X., 2018b. Computer-aided grading of gliomas combining automatic segmentation and radiomics. International Journal of Biomedical Imaging 2018. doi:https://doi.org/10.1155/2018/2512037.

Clough, J.R., Oksuz, I., Puyol-Antón, E., Ruijsink, B., King, A.P., Schnabel, J.A., 2019. Global and local interpretability for cardiac mri classification, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI
2019. Springer International Publishing. Cham. pp. 656–664. doi: 10.1007/978-3-030-32251-9_72

Deng, L., Liu, Y., 2018. Deep Learning in Natural Language Processing. Springer Publishing Company, Incorporated.

Girum, K.B., Skandarani, Y., Hussain, R., Grayeli, A.B., Créhange, G., Lalande, A., 2021. Automatic myocardial infarction evaluation from delayed-enhancement cardiac mri using deep convolutional networks, in: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, Springer International Publishing, Cham. pp. 378–384.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: Teh, Y.W., Titterington, M. (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR, Chia Laguna Resort, Sardinia, Italy. pp. 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., pp. 2672–2680. URL: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Hadjeres, G., Nielsen, F., Pachet, F., 2017. GLSR-VAE: Geodesic latent space regularization for variational autoencoder architectures. 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 1–7. doi: 10.1109/SSCI.2017.8280895

Higaki, T., Nakamura, Y., Zhou, J., Yu, Z., Nemoto, T., Tatsugami, F., Awai, K., 2020. Deep learning reconstruction at ct: Phantom study of the image characteristics. Academic Radiology 27, 82–87. doi: 10.1016/j.acra.2019.09.008

Higgins, I., Matthey, L., Pal, A., Burgess, C.P., Glorot, X., Botvinick, M.M., Mohamed, S., Lerchner, A., 2017a. Beta-vae: Learning basic visual concepts with a constrained variational framework, in: ICLR.

Higgins, I., Sonnerat, N., Matthey, L., Pal, A., Burgess, C.P., Botvinick, M.M., Hassabis, D., Lerchner, A., 2017b. SCAN: Learning abstract hierarchical compositional visual concepts. ArXiv URL: https://arxiv.org/abs/1707.03389

Huang, M.L., Hung, Y.H., Lee, W.M., Li, R.K., Jiang, B.R., 2014. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. ScientificWorldJournal 2014, 795624. doi: 10.1155/2014/795624

Ivantsits, M., Huellebrand, M., Kelle, S., Schönberg, S.O., Kuchne, T., Hennemuth, A., 2021. Deep-learning-based myocardial pathology detection, in: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, Springer International Publishing, Cham. pp. 369–377.

Jo, T., Nho, K., Saykin, A.J., 2019. Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. Frontiers in Aging Neuroscience 11, 220. doi: 10.3389/fnagi.2019.00220

Kapishnikov, A., Bolukbasi, T., Viegas, F., Terry, M., 2019. XRAI: Better attributions through regions. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 4947–4956. doi: 10.1109/ICCV.2019.00505
Kim, H., Mnih, A., 2018. Disentangling by Factorizing, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, PMLR. pp. 2649–2658.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, in: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. URL: https://arxiv.org/abs/1412.6980v9.

Kingma, D.P., Welling, M., 2014. Auto-Encoding Variational Bayes, in: 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings. arXiv:http://arxiv.org/abs/1312.6114v10.

Kofler, A., Haltmeier, M., Schaeffter, T., Kachelriess, M., Dewey, M., Wald, C., Kolbitsch, C., 2020. Neural networks-based regularization for large-scale medical image reconstruction. Physics in medicine and biology 65, 135003. doi:10.1088/1361-6560/ab990e.

Kumar, A., Sattigeri, P., Balakrishnan, A., 2017. Variational inference of disentangled latent concepts from unlabeled observations. CoRR abs/1711.00848. URL: http://arxiv.org/abs/1711.00848. arXiv:1711.00848.

Lalande, A., Chen, Z., Decourselle, T., Qayyum, A., Pommier, T., Lorgis, L., de la Rosa, E., Cochet, A., Cottin, Y., Ginhac, D., Salomon, M., Coudurier, R., Meriaudeau, F., 2020. Emidec: A database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac mri. Data 5. doi:10.3390/data5040089.

Lalande, A., Chen, Z., Pommier, T., Decourselle, T., Qayyum, A., Salomon, M., Ginhac, D., Skandarani, Y., Boucher, A., Brahim, K., de Bruijne, M., Camarasa, R., Correia, T., Feng, X., Girum, K.B., Hennemuth, A., Huellebrand, M., Hussain, R., Ivantsits, M., Ma, J., Meyer, C.H., Sharma, R., Shi, J., Tsekos, N.V., Varela, M., Wang, X., Yang, S., Zhang, H., Zhang, Y., Zhou, Y., Zhuang, X., Coudurier, R., Mériaudeau, F., 2021. Deep learning methods for automatic evaluation of delayed enhancement-mri: the results of the emidec challenge. CoRR abs/2108.04016. URL: https://arxiv.org/abs/2108.04016. arXiv:2108.04016.

Lambin, P., Leijenaar, R.T.H., Deist, T.M., Peerlings, J., de Jong, E.E.C., van Timmeren, J., Sanduleanu, S., Larue, R.T.H.M., Even, A.J.G., Jochems, A., van Wijk, Y., Woodruff, H., van Soest, J., Lustberg, T., Roelofs, E., van Elmpt, W., Dekker, A., Mottaghy, F.M., Wildberger, J.E., Walsh, S., 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14, 749–762. doi:10.1038/nrclinonc.2017.141.

Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., Ranzato, M., 2017. Fader networks: Manipulating images by sliding attributes, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA. p. 5969–5978.

Larroza, A., López-Lereu, M.P., Monmeneu, J.V., Gavara, J., Chorro, F.J., Bodi, V., Moratal, D., 2018. Texture analysis of cardiac cine magnetic resonance imaging to detect nonviable segments in patients with chronic myocardial infarction. Medical Physics 45, 1471–1480. doi:10.1002/mp.12783.

Liu, W., Li, R., Zheng, M., Karanam, S., Wu, Z., Bhanu, B., Radke, R.J., Camps, O.I., 2020. Towards visually explaining variational autoencoders, pp. 8639–8648. doi:10.1109/CVPR42600.2020.00867.

Liu, X., Sanchez, P., Thermos, S., O’Neil, A.Q., Tsafaritis, S.A., 2021. Learning disentangled representations in the imaging domain. URL: https://arxiv.org/abs/2108.12043. arXiv:2108.12043.

Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., Bachem, O., 2019. Challenging common assumptions in the unsupervised learning of disentangled representations, in: Chaudhuri, K., Salakhutdinov, R. (Eds.), Proceedings of the 36th International Conference on Machine Learning, PMLR. pp. 4114–4124. URL: https://proceedings.mlr.press/v97/locatello19a.html.

López-Linares, K., Aranjuelo, N., Kabongo, L., Maclaur, G., Lete, N., Ceresa, M., García-Familiar, A., Macía,
I., González Ballester, M.A., 2018. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative cta images using deep convolutional neural networks. Medical Image Analysis 46, 202–214. doi:10.1016/j.media.2018.03.010

Lourenço, A., Kerfoot, E., Grigorescu, I., Scannell, C.M., Varela, M., Correia, T.M., 2021. Automatic myocardial disease prediction from delayed-enhancement cardiac mri and clinical information, in: Puyol Antón, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, Springer International Publishing, Cham. pp. 334–341.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., pp. 4768–4777.

Painchaud, N., Duchateau, N., Bernard, O., Jodoin, P.M., 2021. Echocardiography segmentation with enforced temporal consistency. ArXiv URL: https://arxiv.org/pdf/2112.02102.pdf

Pati, A., Lerch, A., 2021. Attribute-based Regularization of Latent Spaces for Variational Auto-Encoders. Neural Computing and Applications 33, 4429—4444. doi:10.1007/s00521-020-05270-2

Puyol-Antón, E., Chen, C., Clough, J.R., Ruijsink, B., Sidhu, B.S., Gould, J., Porter, B., Elliott, M., Mehta, V., Rueckert, D., Rinaldi, C.A., King, A.P., 2020. Interpretable deep models for cardiac resynchronisation therapy response prediction, 284–293 doi:10.1007/978-3-030-59710-8_28

Reyes, M., Meier, R., Pereira, S., Silva, C.A., Dahlweid, F.M., von Tengg-Kobligk, H., Summers, R.M., Wiest, R., 2020. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. Radiol Artif Intell 2, e190043. doi:10.1148/ryai.2020190043

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. “Why Should I Trust You?”: Explaining the predictions of any classifier, in: Proceedings of the Demonstrations Session, NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016. The Association for Computational Linguistics. pp. 97–101. doi:10.18653/v1/n16-3020

Ridgeway, K., Mozer, M.C., 2018. Learning deep disentangled embeddings with the f-statistic loss, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., pp. 185–194.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convoluted networks for biomedical image segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham. pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Rubenstein, P.K., Schölkopf, B., Tolstikhin, I.O., 2018. Learning disentangled representations with wasserstein auto-encoders, in: ICLR.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 206–215. URL: https://doi.org/10.1038/s42256-019-0048-x, doi:10.1038/s42256-019-0048-x.

Salahuddin, Z., Woodruff, H.C., Chatterjee, A., Lambin, P., 2022. Transparency of deep neural networks for medical image analysis: A review of interpretability methods. Computers in Biology and Medicine 140. URL: https://www.sciencedirect.com/science/article/pii/S0010482521009057, doi:10.1016/j.compbiomed.2021.105111.

Samb, M.L., Camara, F., Ndiaye, S., Slimani, Y., Esseghir, M.A., Anta, C., 2012. A novel rfe-svm-based feature selection approach for classification. International Journal of Advanced Science and Technology 43. doi:10.1155/2014/795624.

Sanchez-Martinez, S., Camara, O., Piella, G., Cikes, M., González Ballester, M.A., Miron, M., V.A., Gomez, E., Fraser, A., Bijnen, B., 2022. Machine learning for clinical decision-making: Challenges and opportunities in Cardiovascular imaging. Frontiers in Cardiovascular Medicine 8. doi:10.3389/fcvm.2021.765693.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626. doi:10.1109/ICCV.2017.74.

Shakiri, M., Lombaert, H., Tripathi, S., Kadoury, S., 2016. Deep spectral-based shape features for alzheimer’s disease classification, in: Reuter, M., Wachinger, C., Lombaert, H. (Eds.), Spectral and Shape Analysis in Medical Imaging. Springer International Publishing, Cham. pp. 15–24. doi:http://dx.doi.org/10.1007/978-3-319-51237-2_2.

Sharma, R., Eick, C.F., Tsekos, N.V., 2021. Sm2n2: A stacked architecture for multimodal data and its application to myocardial infarction detection, in: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, Springer International Publishing, Cham. pp. 342–350.

Shen, D., Wu, G., Suk, H.I., 2017. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 19, 221–248. doi:10.1146/annurev-bioeng-071516-044422.

Shi, J., Chen, Z., Couturier, R., 2021. Classification of pathological cases of myocardial infarction using convolutional neural network and random forest, in: Puyol Anton, E., Pop, M., Sermesant, M., Campello, V., Lalande, A., Lekadir, K., Suinesiaputra, A., Camara, O., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, Springer International Publishing, Cham. pp. 406–413.

Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps, in: Bengio, Y., LeCun, Y. (Eds.), 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings. URL: http://arxiv.org/abs/1312.6034.

Singh, A., Sengupta, S., Lakshminarayanan, V., 2020. Explainable deep learning models in medical image analysis. Journal of Imaging 6. URL: https://www.mdpi.com/2313-433X/6/6/52, doi:10.3390/jimaging6060052.

Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., Soni, S., Wang, Q., Wei, Q., Xiang, Y., Zhao, B., Xu, H., 2019. Deep learning in clinical natural language processing: a methodical review. Journal of the
American Medical Informatics Association 27, 457–470. doi:10.1093/jamia/ocz200.

Xiao, G., Rong, W.C., Hu, Y.C., Shi, Z.Q., Yang, Y., Ren, J.L., Cui, G.B., 2020. MRI Radiomics Analysis for Predicting the Pathologic Classification and TNM Staging of Thymic Epithelial Tumors: A Pilot Study. AJR Am J Roentgenol 214, 328–340. doi:10.2214/AJR.19.21696.

Yang, X., 2018. Identification of risk genes associated with myocardial infarction based on the recursive feature elimination algorithm and support vector machine classifier. Mol Med Rep 17, 1555–1560. doi:10.3892/mmr.2017.8044.

Zhang, X., Xu, X., Tian, Q., Li, B., Wu, Y., Yang, Z., Liang, Z., Liu, Y., Cui, G., Lu, H., 2017. Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging. J Magn Reson Imaging 46, 1281–1288. doi:10.1002/jmri.25669.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. doi:10.1109/ICCV.2017.244.