Detecting Unforeseen Data Properties with Diffusion Autoencoder Embeddings using Spine MRI data

Funded by
the European Union

Robert Graf, Florian Hunecke, Soeren Pohl, Matan Atad, Hendrik Moeller, Sophie Starck, Thomas Kroencke, Stefanie Bette, Fabian Bamberg, Tobias Pischon, Thoralf Niendorf, Carsten Schmidt, Johannes C. Paetzold, Daniel Rueckert, and Jan S Kirschke

What are embeddings?

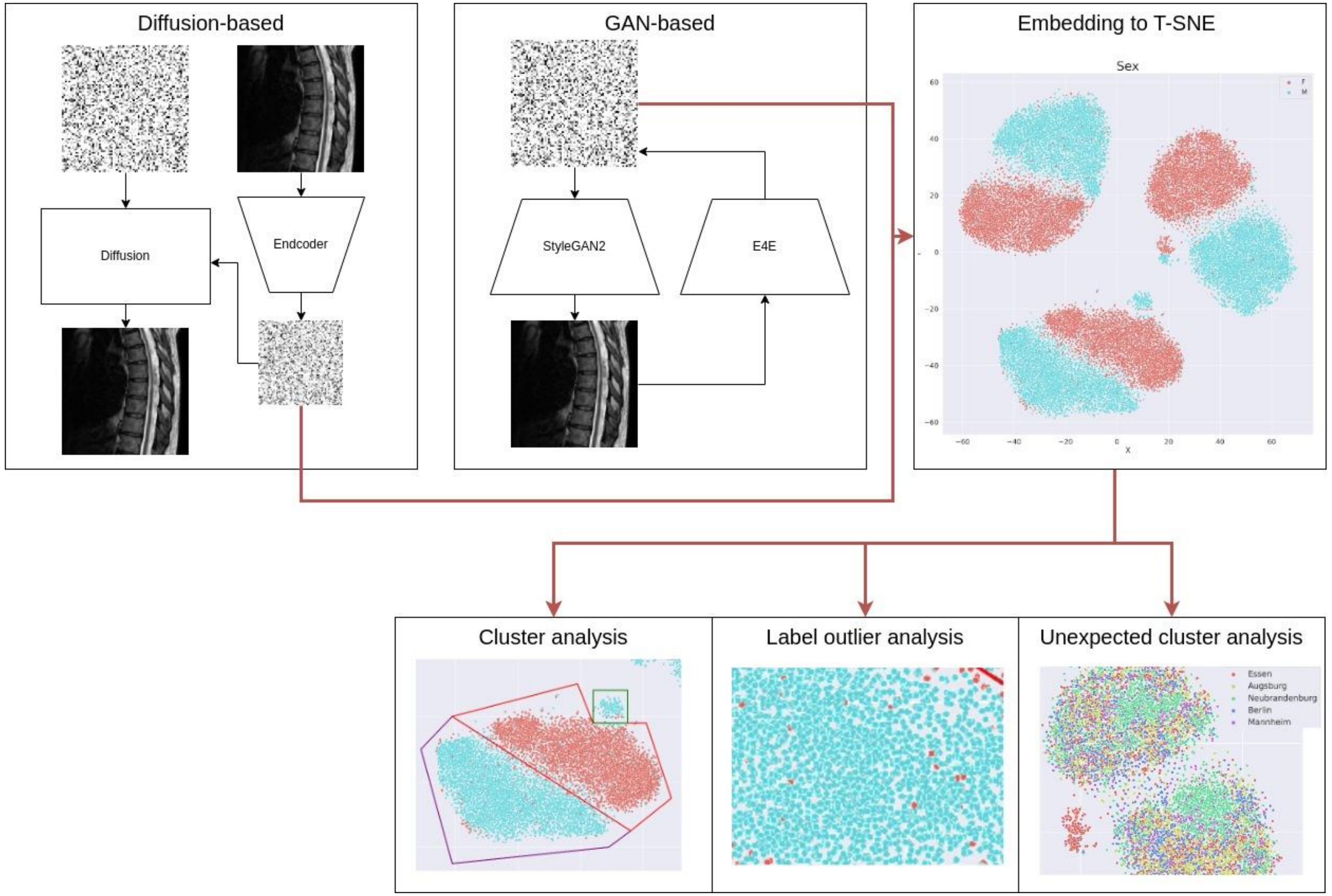
Embeddings are numerical representations of data, such as words or images, in a lower-dimensional space. They capture the relationships and similarities between data points.
~chatGPT

Unsupervised

Generative

Latents from a supervised Task

Language modeling



Why generative embedding for explainability?

Your data bias will become your model bias!

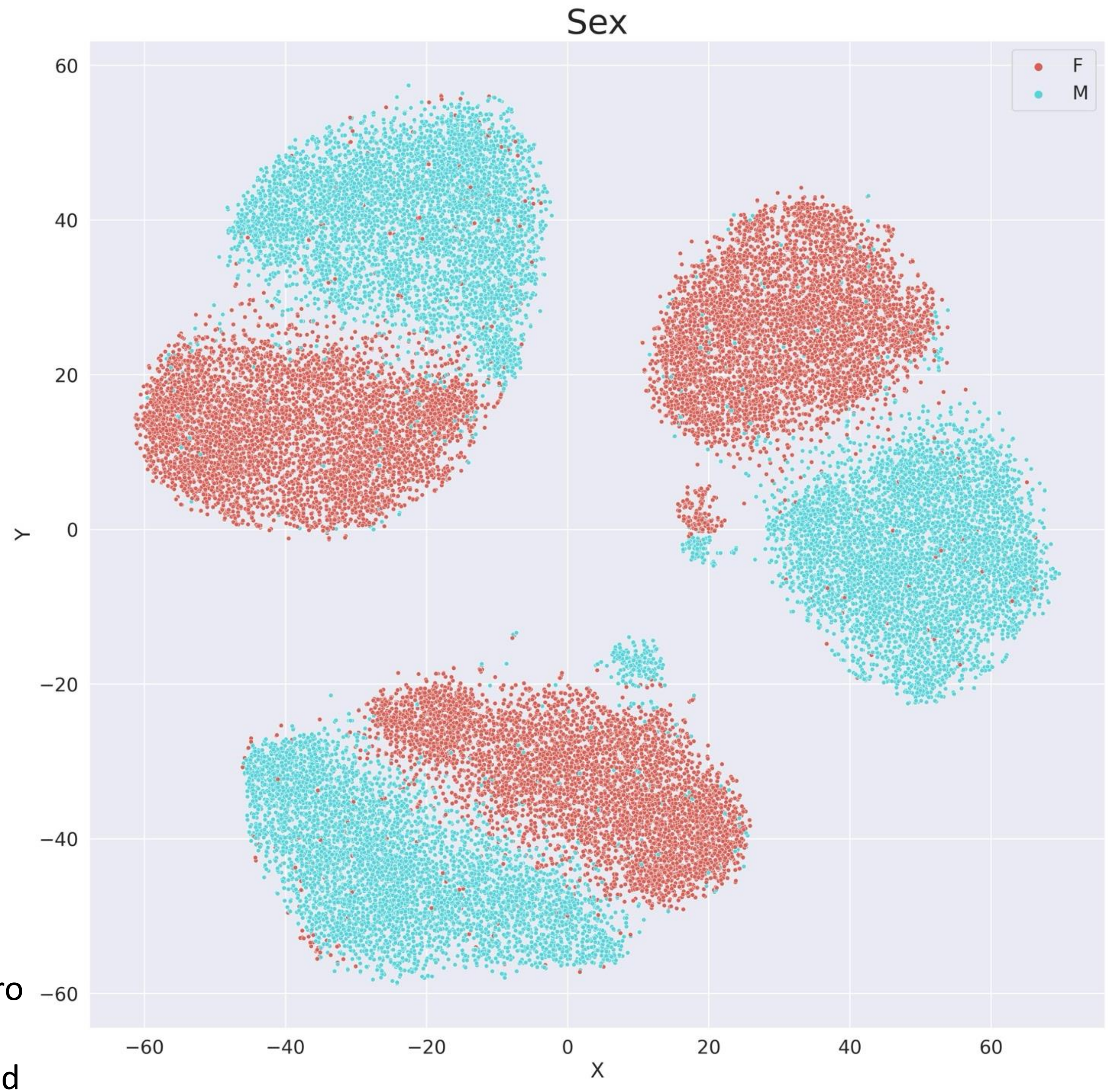
Generative embeddings are independent of your labels and show:
- Natural Bias
- Selection Bias
- Possible areas where learning/labeling is difficult.

We can use Embedding interpolation to combat existing data imbalances and unfairness. [1,2]

[1] SMOTE: Synthetic Minority Over-sampling Technique (https://arxiv.org/abs/1106.1813)
[2] SMOTified-GAN for class imbalanced pattern classification problems (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9733348)

Method

1. Train a DAE[1,2]
   or StyleGAN2 [3]
   or beta-VAE [4]
2. Compute embeddings
3. Use T-SNE plots for down-projection
4. Color the points by labels



DAE embeddings color by Sex

[1] Diffusion Autoencoders: Toward a Meaningful and Decodable Representation, Konpat Preechakul et Al
[2] Pluralistic Aging Diffusion Autoencoder, Peipei Li et Al
[3]Analyzing and Improving the Image Quality of StyleGAN, Tero Karras et Al
[4] beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, Irina Higgins et Al

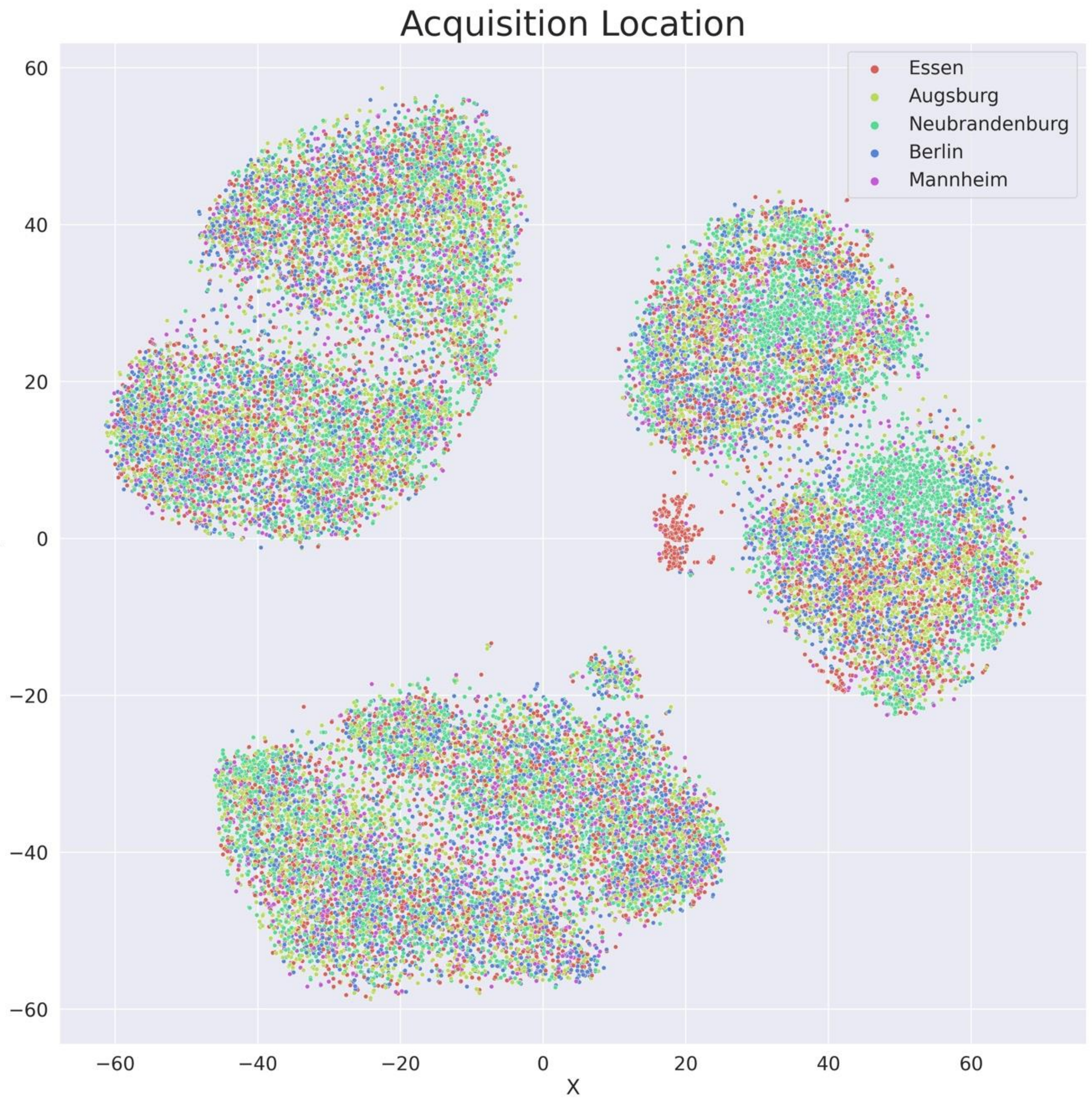
Method

1. Train a DAE[1,2]
   or StyleGAN2 [3]
   or beta-VAE [4]
2. Compute embeddings
3. Use T-SNE plots for down-projection
4. Color the points by labels

Found something unexpected?

1. Delineate those Clusters
2. Run a classifier on these delineation
3. Evaluate with a GradCAM variant

[1] Diffusion Autoencoders: Toward a Meaningful and Decodable Representation, Konpat Preechakul et Al
[2] Pluralistic Aging Diffusion Autoencoder, Peipei Li et Al
[3]Analyzing and Improving the Image Quality of StyleGAN, Tero Karras et Al
[4] beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, Irina Higgins et Al



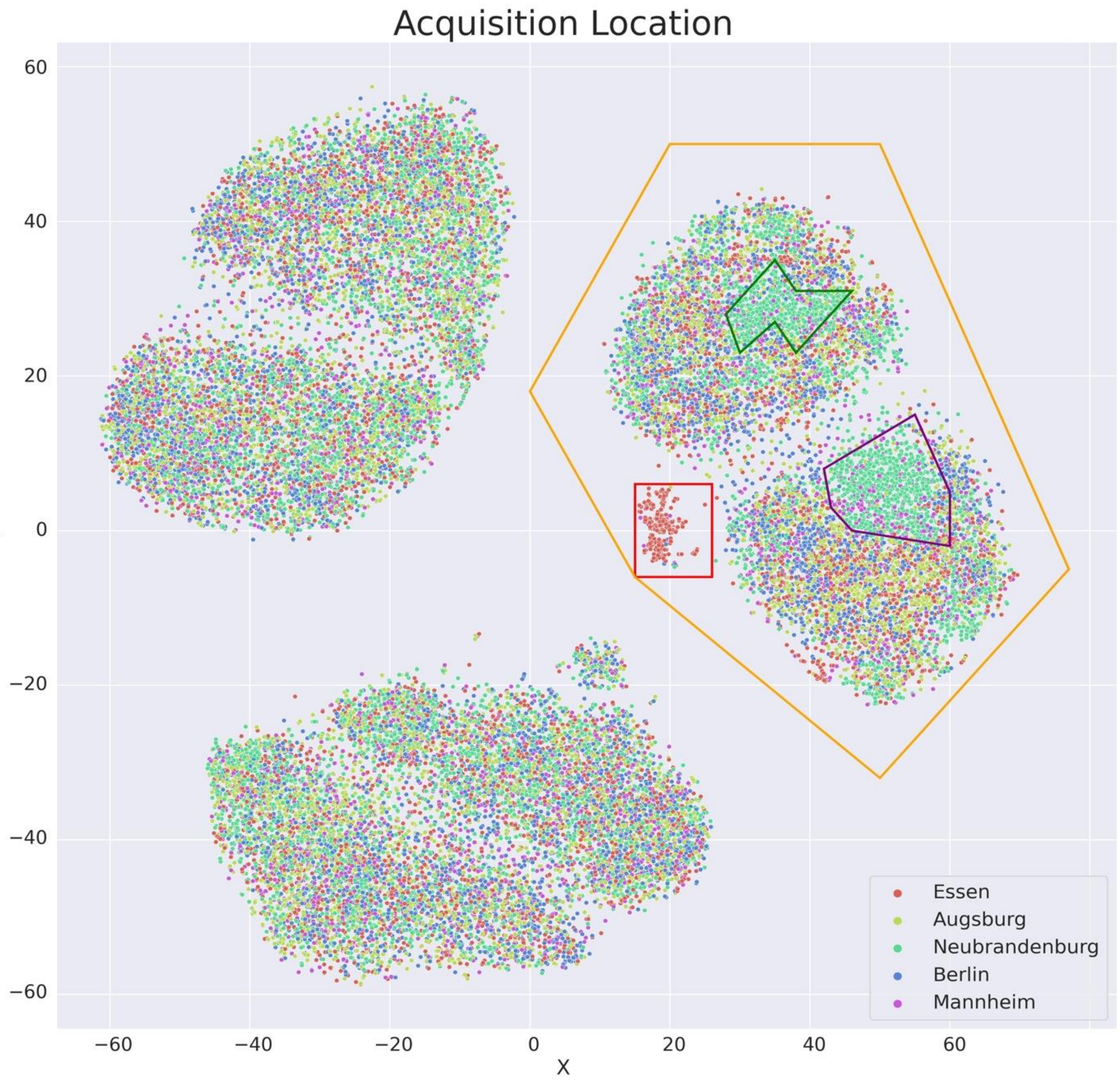
DAE embeddings colored where the MR was acquired

Method

1. Train a DAE[1,2]
   or StyleGAN2 [3]
   or beta-VAE [4]
2. Compute embeddings
3. Use T-SNE plots for down-projection
4. Color the points by labels

Found something unexpected?

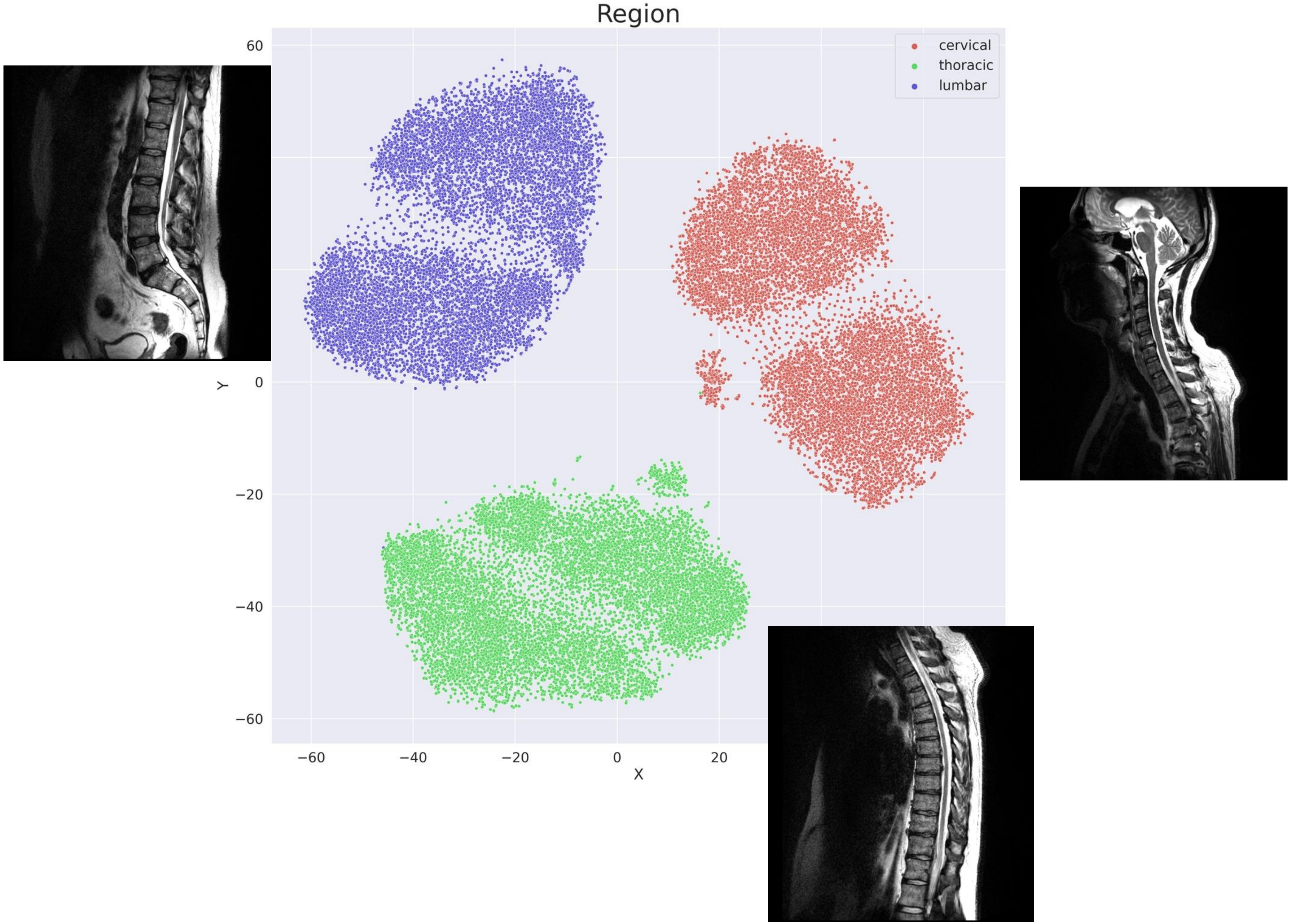
1. Delineate those Clusters
2. Run a classifier on these delineation
3. Evaluate with a GradCAM variant



DAE embeddings colored where the MR was acquired
We should not see this clusters.

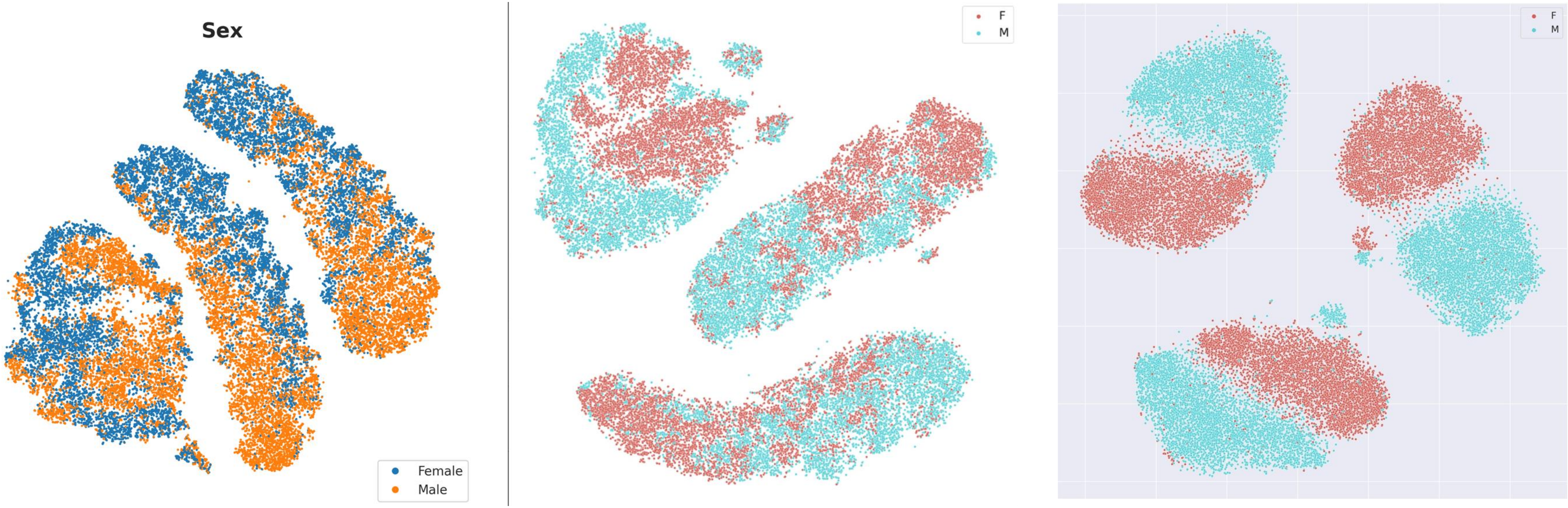
[1] Diffusion Autoencoders: Toward a Meaningful and Decodable Representation, Konpat Preechakul et Al
[2] Pluralistic Aging Diffusion Autoencoder, Peipei Li et Al
[3]Analyzing and Improving the Image Quality of StyleGAN, Tero Karras et Al
[4] beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, Irina Higgins et Al

What are the three clusters?



Results – DAE produce better embeddings than StyleGAN2 and VAEs

Beta-VEA | StyleGAN2 | DAE



Different Clustering of the same data with Beta-VAE, StyleGAN2 and DAE. DAE is able to separate the sex for most subjects.

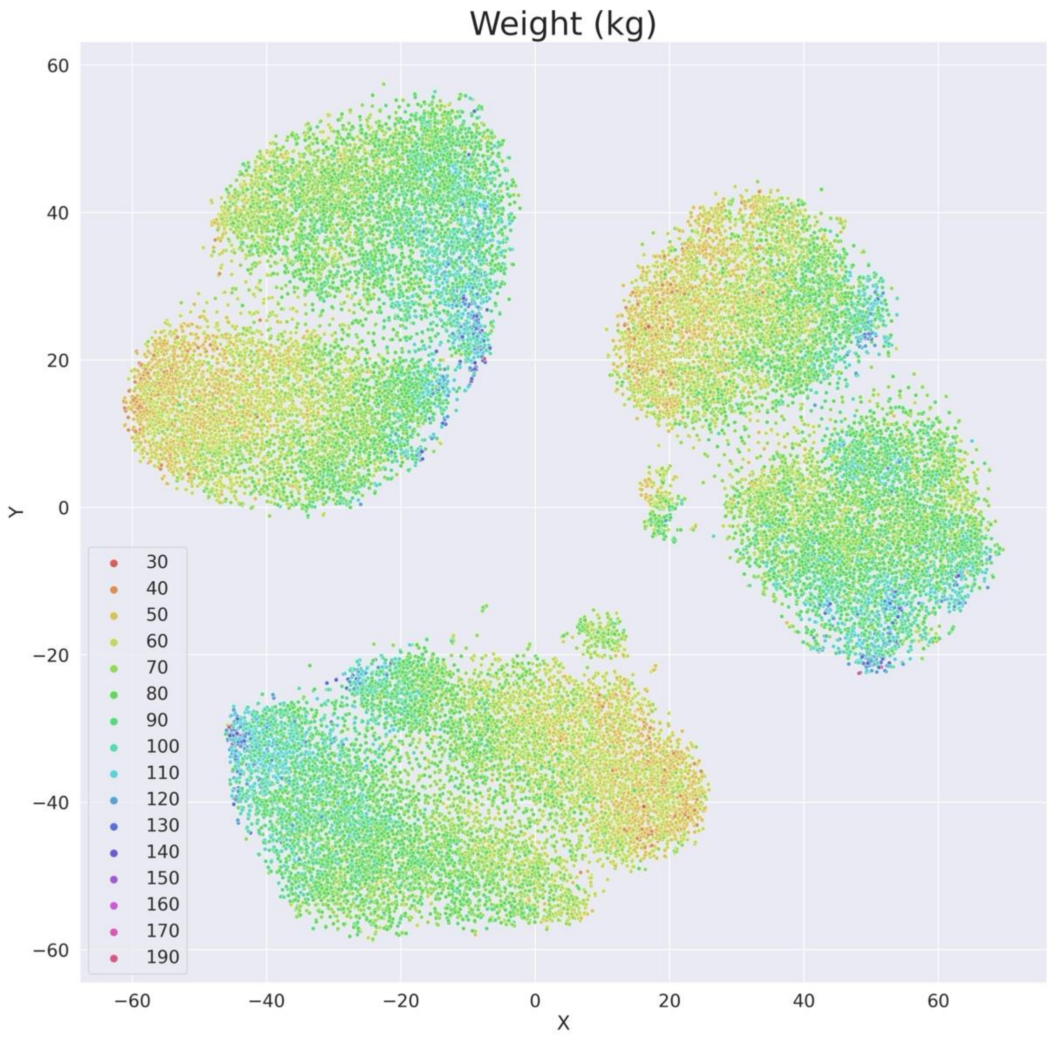
Results – DAE produce better embeddings than StyleGAN2 and VAEs

Table 1. Regression and classification with images and embedding.

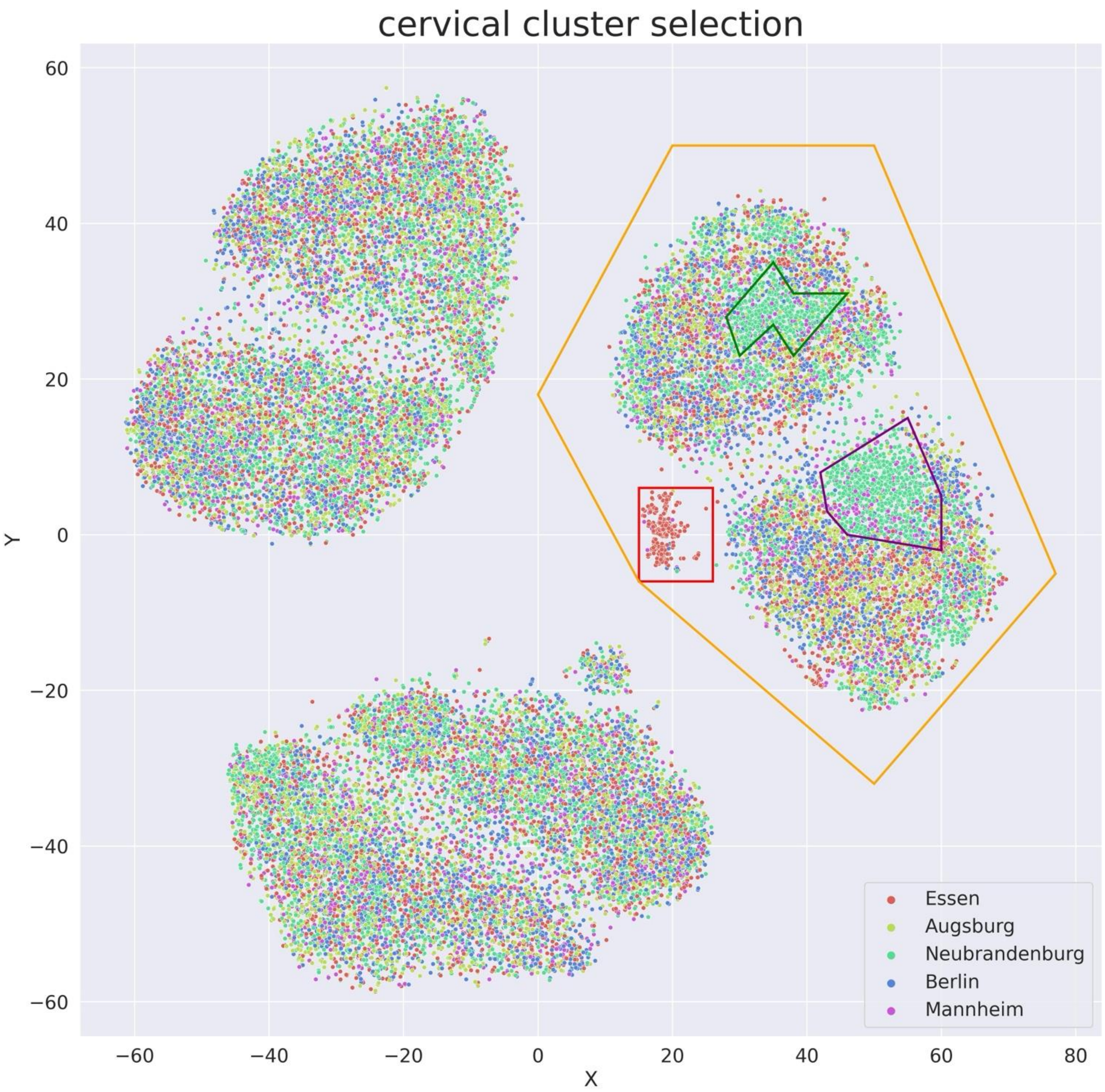
	Super-vision	Type	Body Region accuracy ↑	Sex accuracy ↑	Weight ℓ_1 kg ↓	Height ℓ_1 meter ↓	Age ℓ_1 years ↓
β-VAE + Hessian	semi	embed.	0.998	0.870	6.75	0.055	7.80
StyleGAN	semi	embed.	**1.000**	0.885	5.66	0.047	5.62
DAE (ours)	semi	embed.	**1.000**	**0.988**	**4.32**	**0.032**	**3.84**
ResNet10	fully	image	0.997	0.993	10.26	0.072	4.15

Training an SVM on the embeddings.

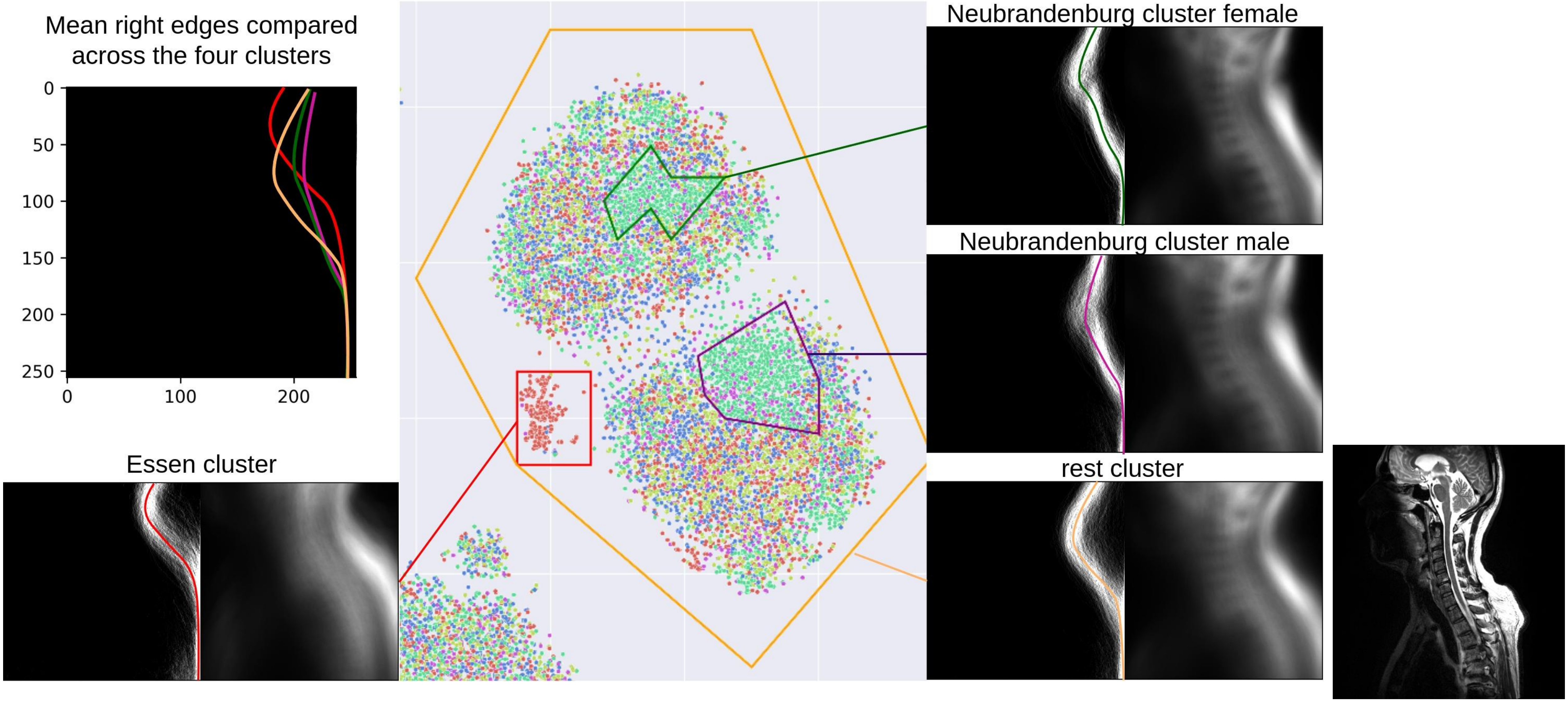
Results – DAE embeddings



Results – DAE embeddings

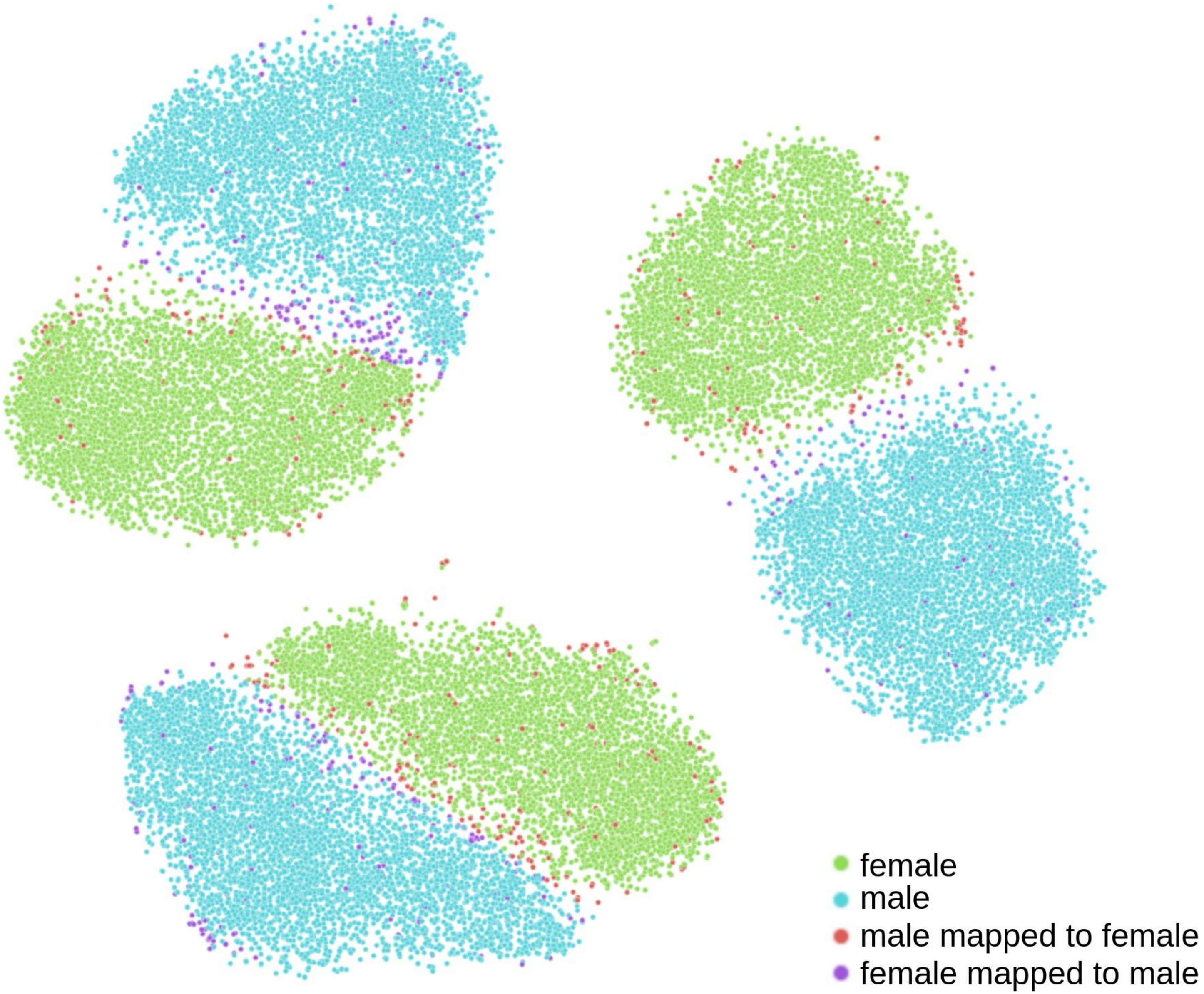


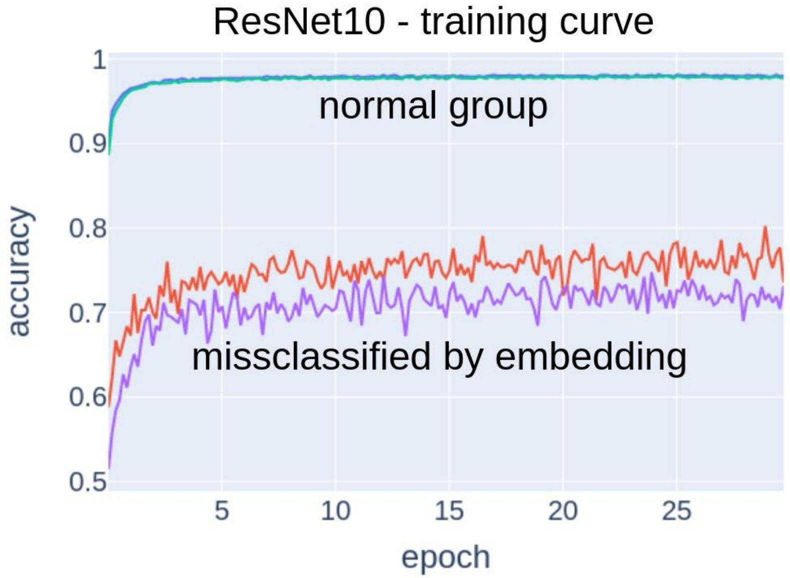
Results – DAE embeddings Acquisition Location

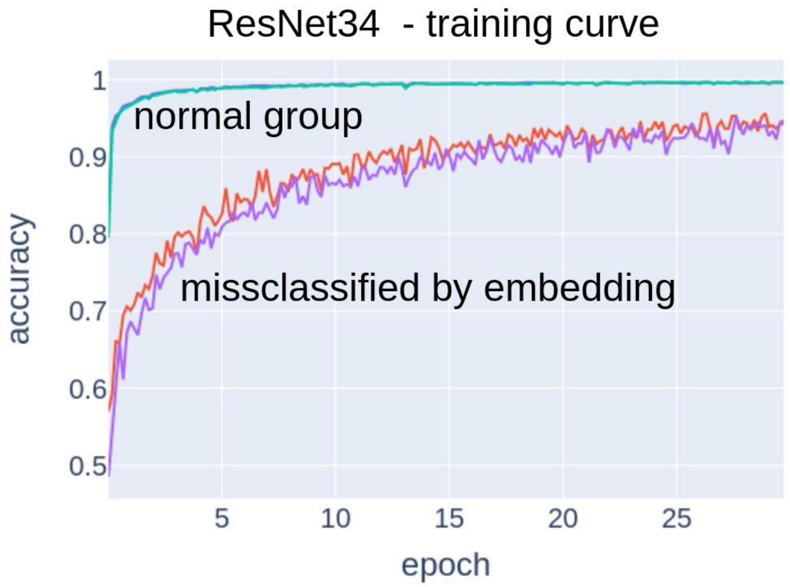


DAE embeddings colored where the MR was acquired. Clusters show that head images in essen are shifted by 50 voxel and Neubrandenburg/Mannheim have different headrests.

Results – DAE embeddings Sex



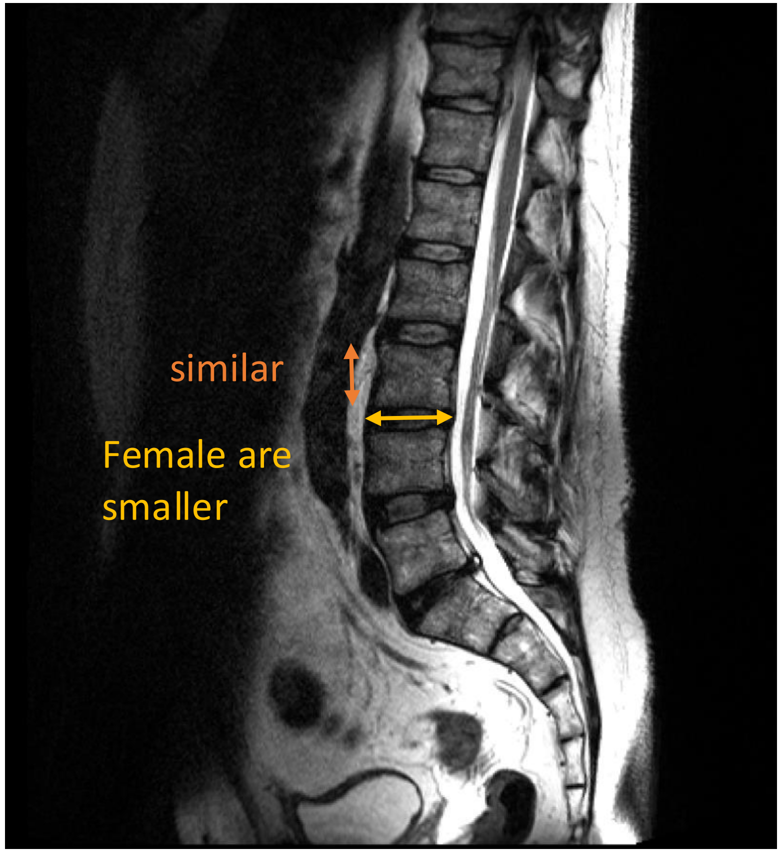




Some individuals have a sex label that does not align with the anatomical expression of their spine. These cases can only be learned through memorization.

Sex differences

The GradCAM highlighted areas around the vertebra as part of its explanation.



Men and women have similar vertebra height relative to their size, but men generally have a much wider vertebra.

Future

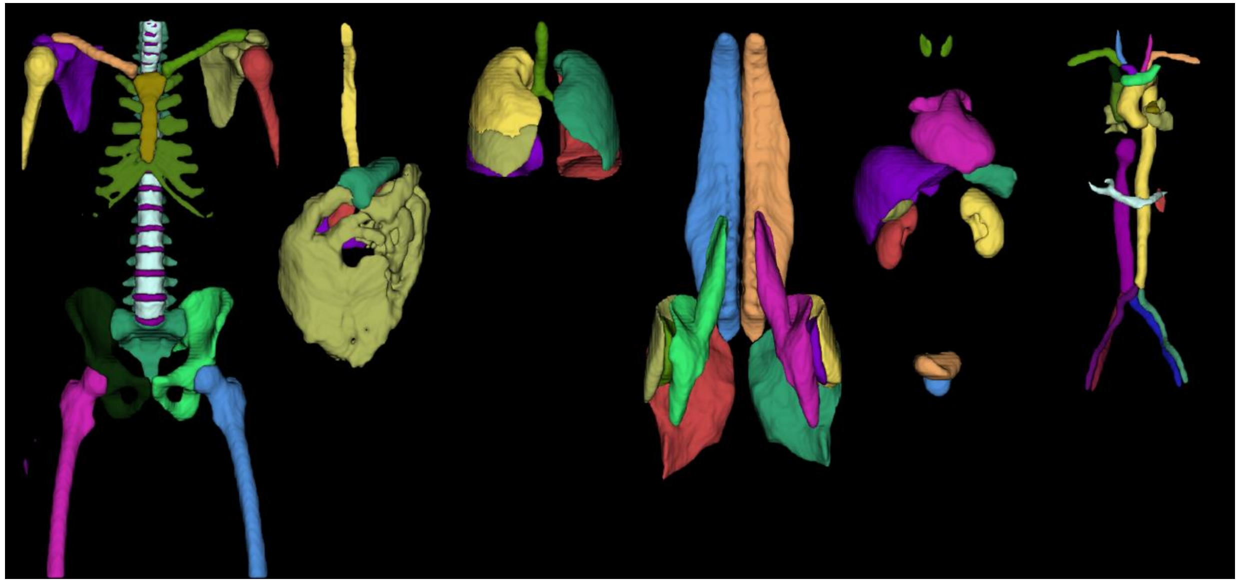
From Global to Local Embeddings

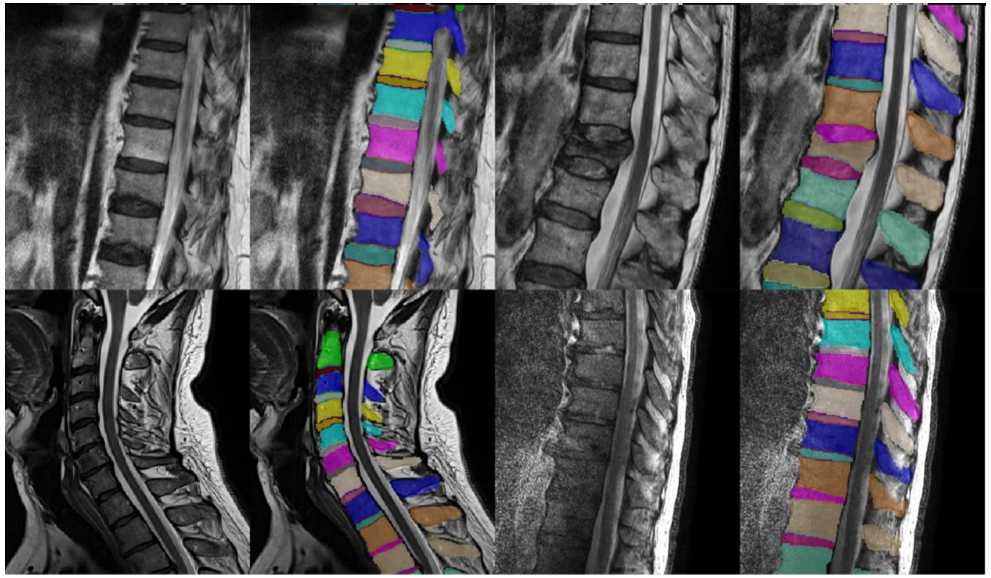
- Utilize MR segmentation to restrict views.
- Clustering desise by embedding look up [1]

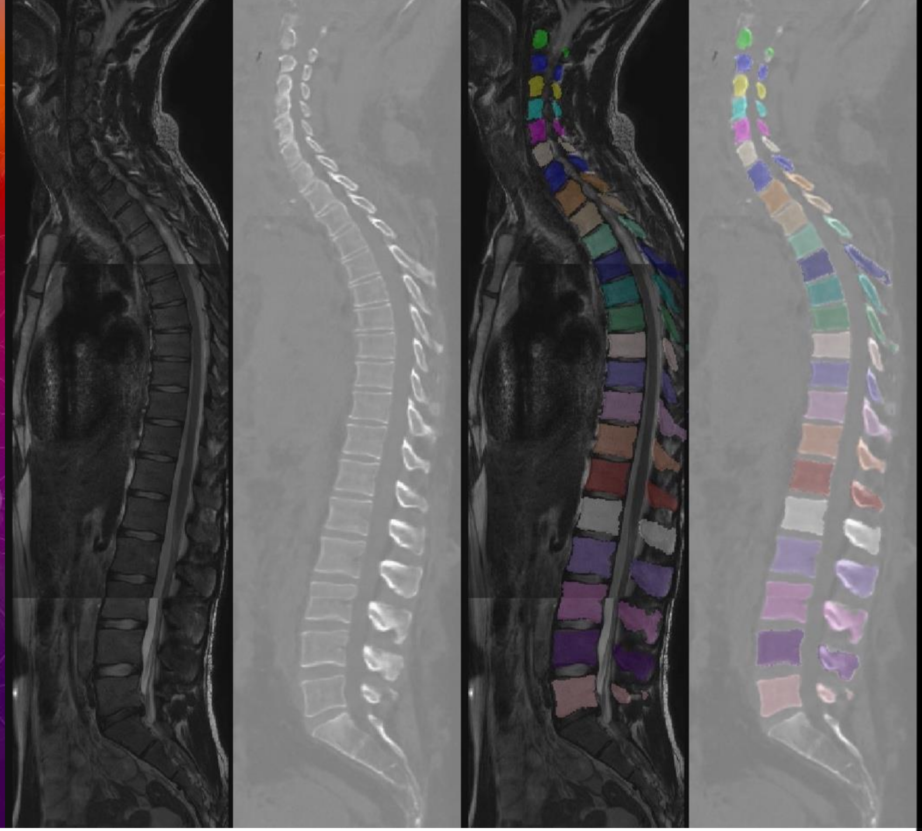
Embedding Interpolation as a Method for Data Imbalance Mitigation

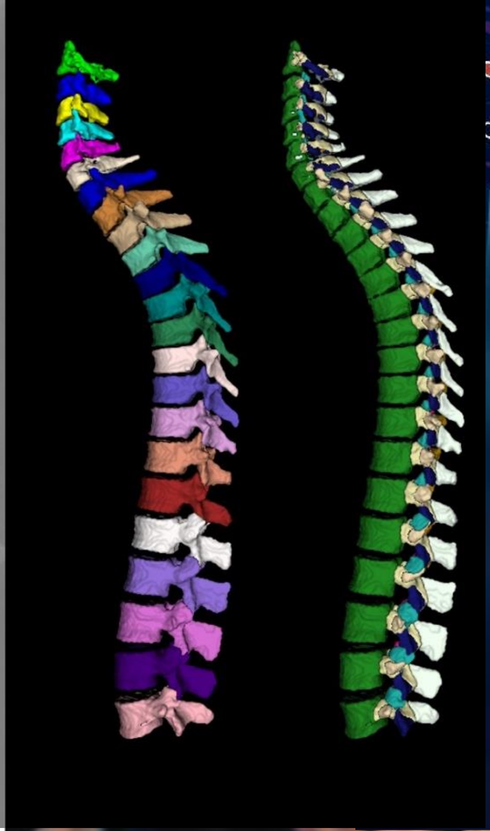
- Can this reduce known biases and increase fairness?
- Can we preserve disease characteristics while modifying epidemiological values?

[1] Hinterwimmer, F., Serena, R.S., Wilhelm, N. *et al.* Recommender-based bone tumour classification with radiographs—a link to the past. *Eur Radiol* **34**, 6629–6638 (2024). https://doi.org/10.1007/s00330-024-10672-0









Thanks for your Attention

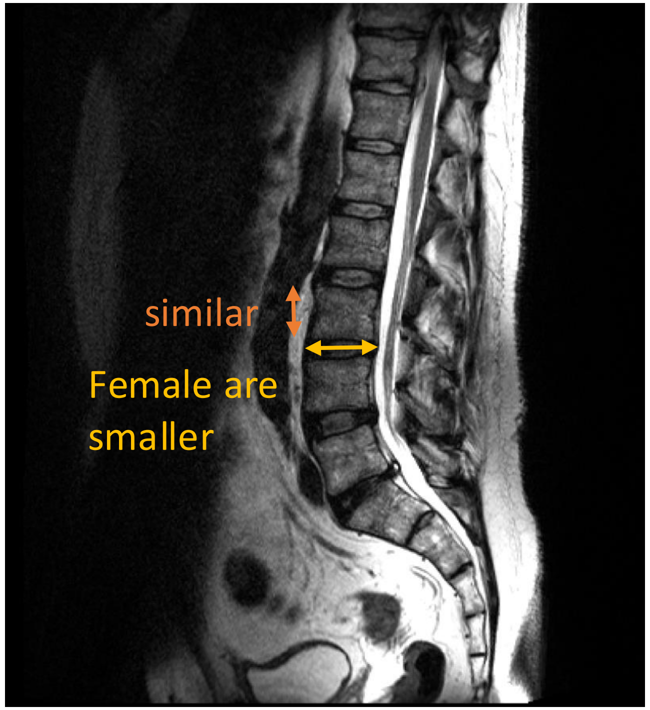
iBack-epic

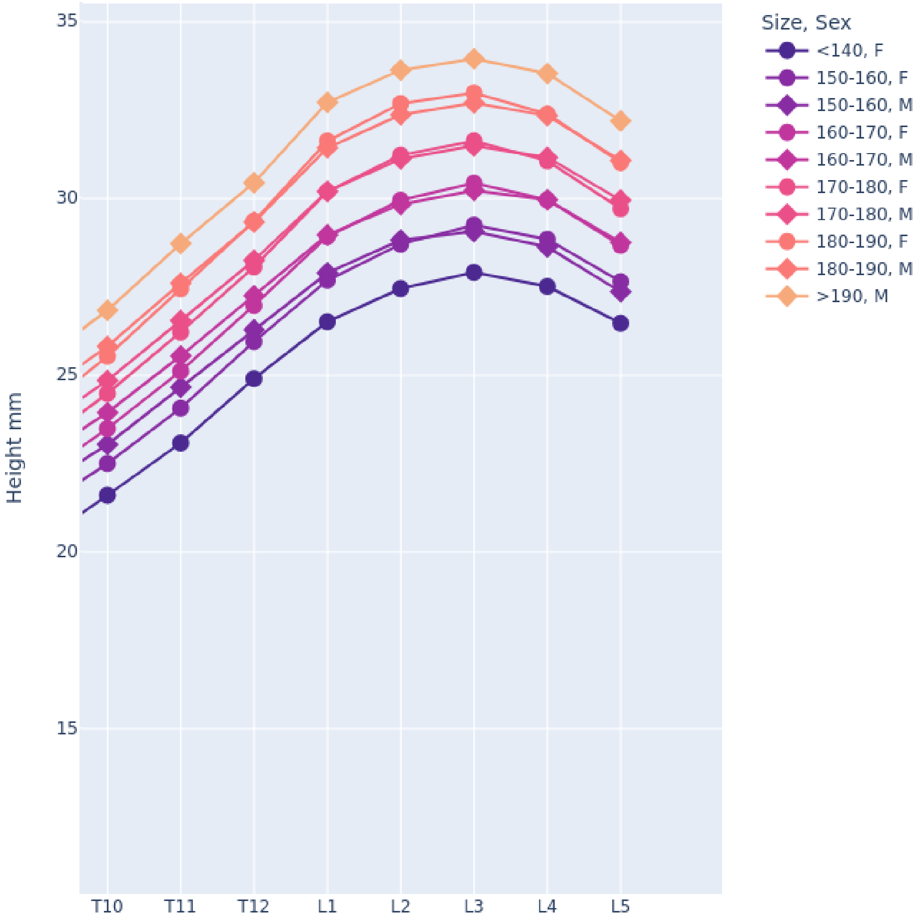
Funded by
the European Union

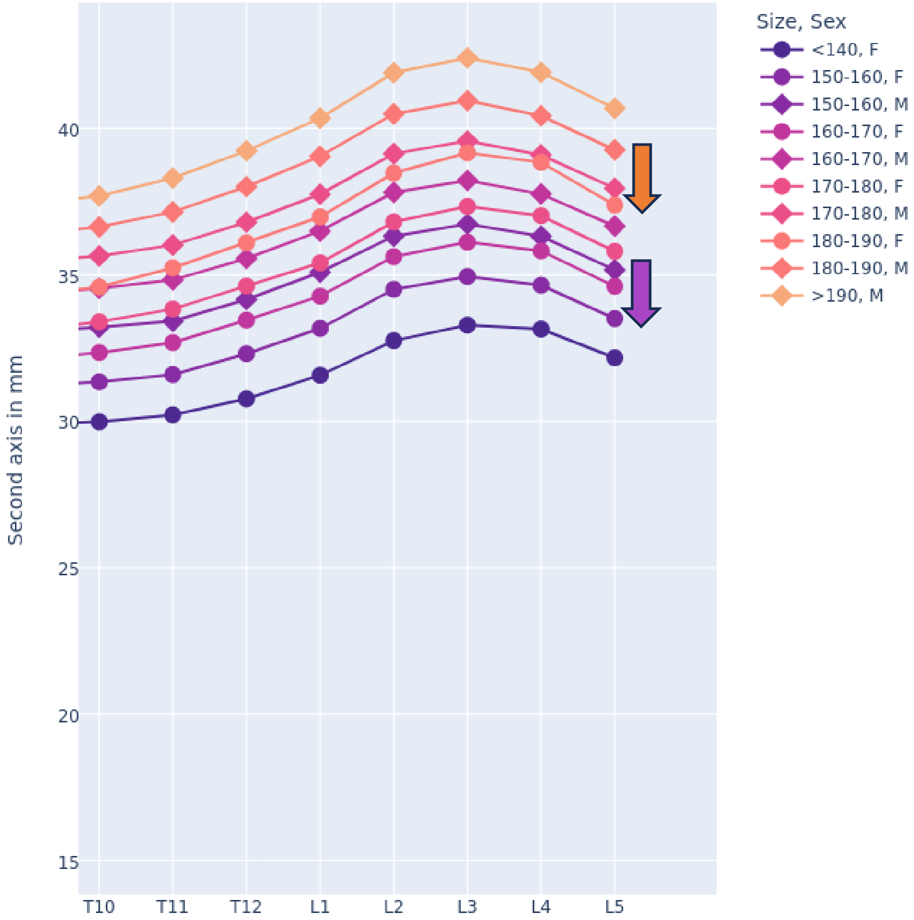
Sex differences

Men and women have similar vertebra height relative to their size, but men generally have a much wider vertebra.

The GradCAM highlighted areas around the vertebra as part of its explanation.

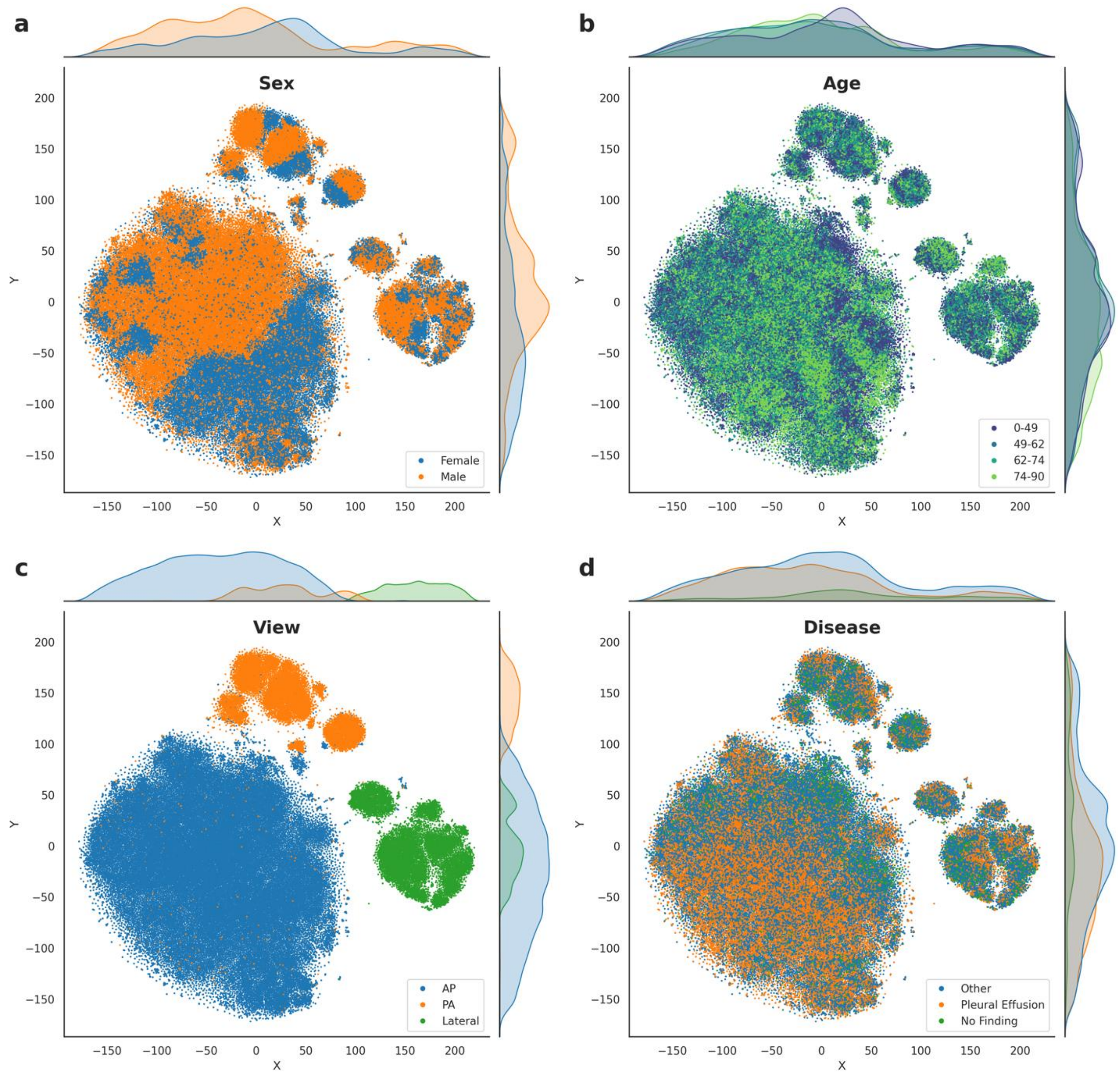


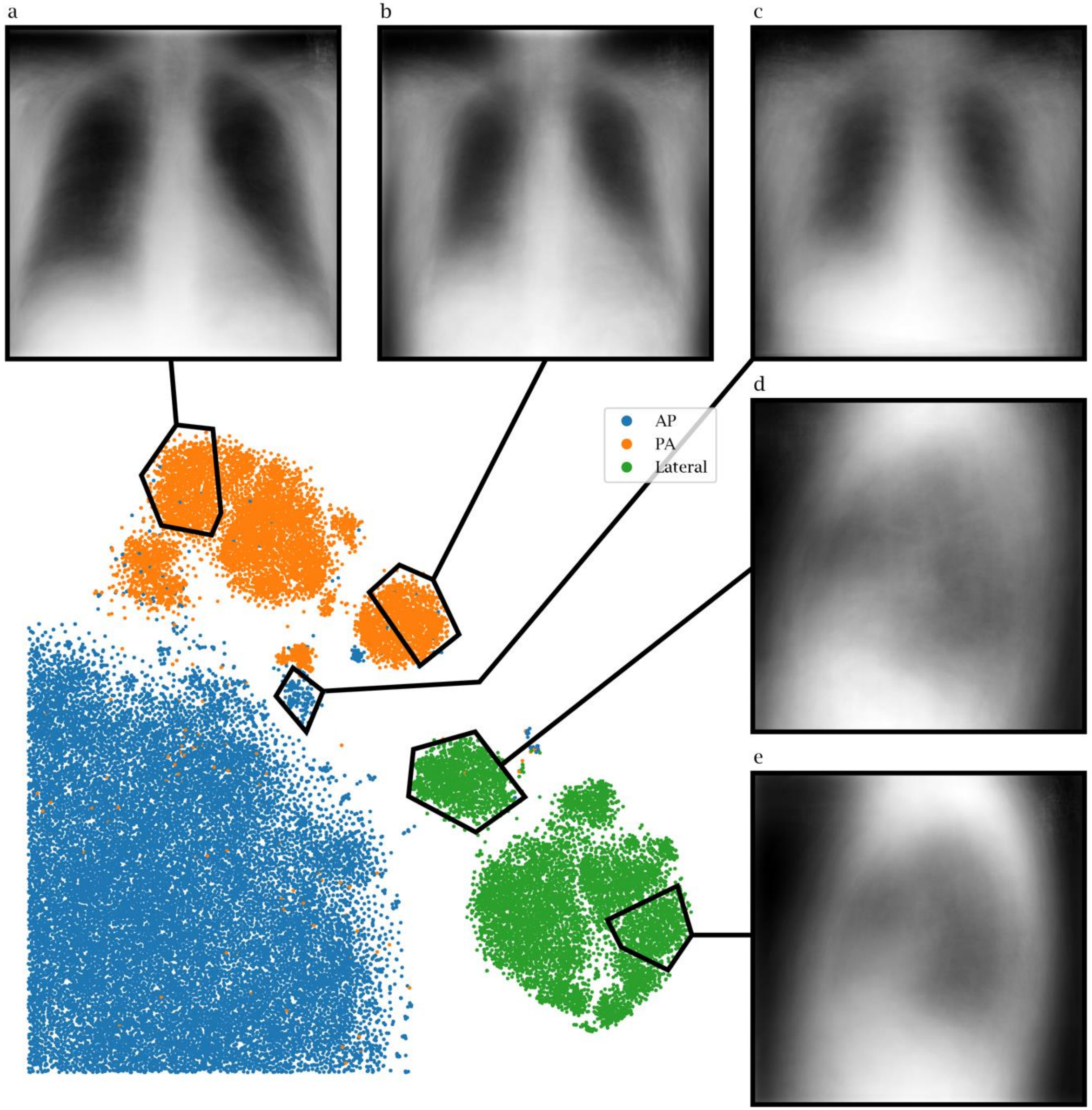




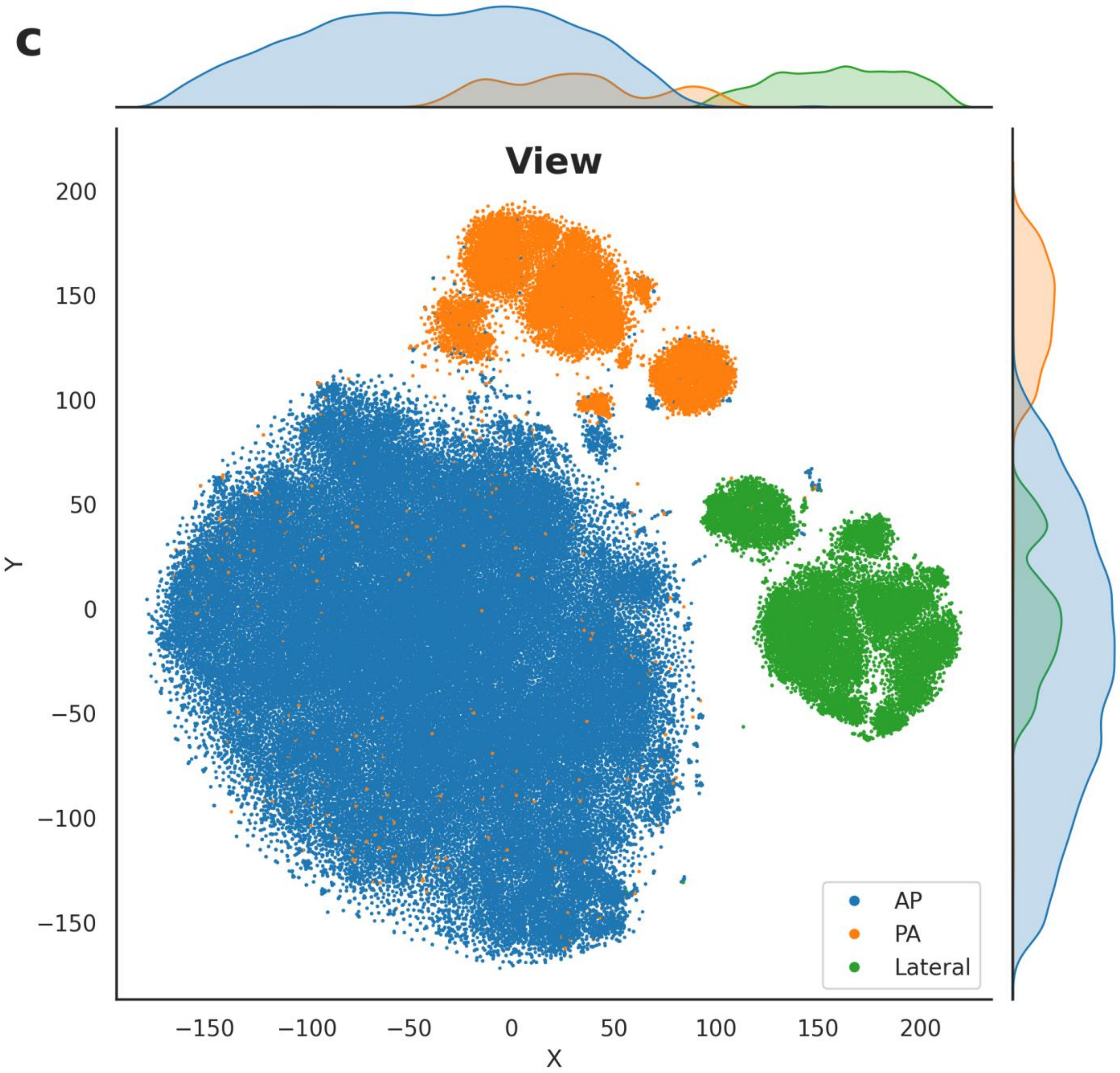
Vertebra height and width analyzed from 30,000 T2w images in the NAKO dataset. Preliminary results for norm value extraction.

Reproduction on Chexpert





Reproduction on Chexpert



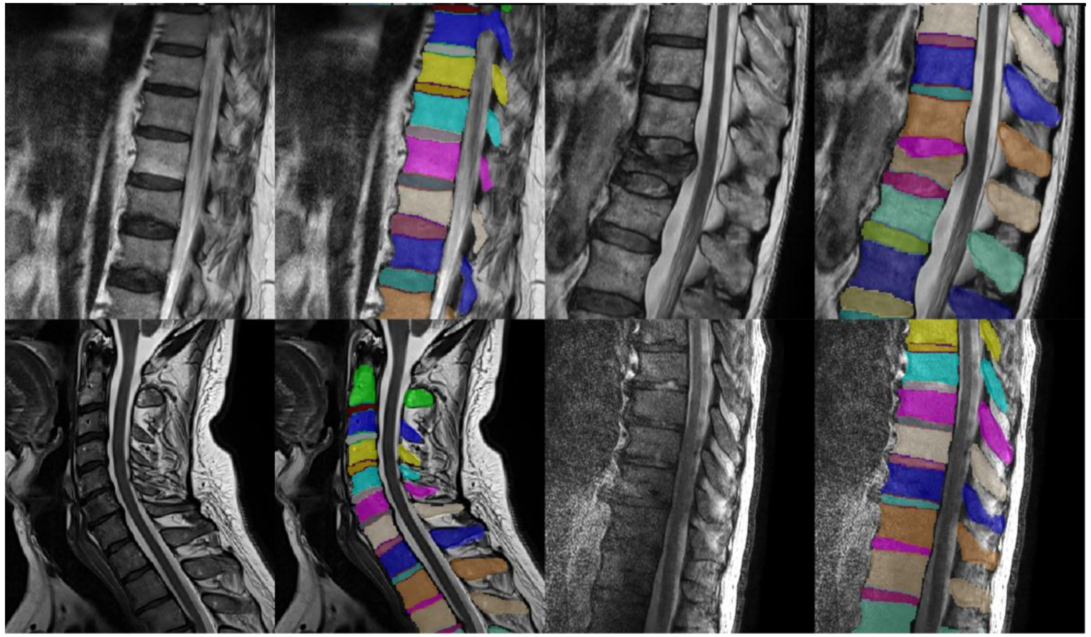
Future

From Global to Local Embeddings

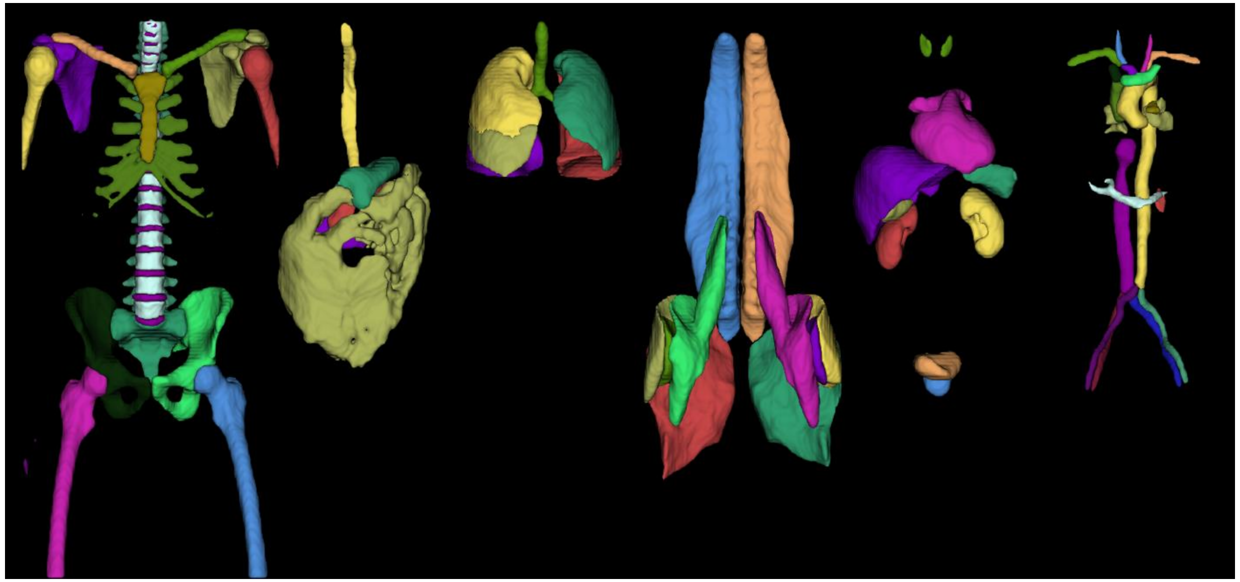
Utilize MR segmentation to restrict views.

Embedding Interpolation as a Method for Data Imbalance Mitigation

- Can this reduce known biases and increase fairness?
- Can we preserve disease characteristics while modifying epidemiological values



Spineps



TotalVibe Segmentator