ETH zürich

Benedikt Dietz
August 9th, 2019

Learning to Diagnose Diabetes from Magnetic Resonance Tomography

Benedikt Dietz
Master Thesis – Final Presentation

Supervision: Patrick Schwab
Dr. Stefan Bauer

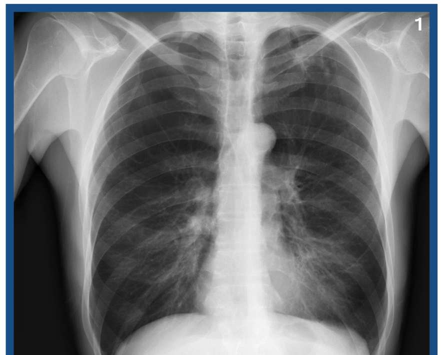
Max-Planck-Institute für
Intelligente Systeme

UNIVERSITÄTS KLINIKUM TÜBINGEN

Introduction & Motivation

High Potential of Machine Learning in Medicine

Modern medicine collects large amounts of unstructured data



Input
Chest X-Ray Image

Machine Learning →

Output
Pneumonia Positive (85%)

Chest X–Rays → **Pneumonia**[1]

Brain MRT → **Alzheimers**[2]

Chest CT → **Lung Cancer**[3]

Retinal Fundus Images → **Diabetic Retinopathy**[4]

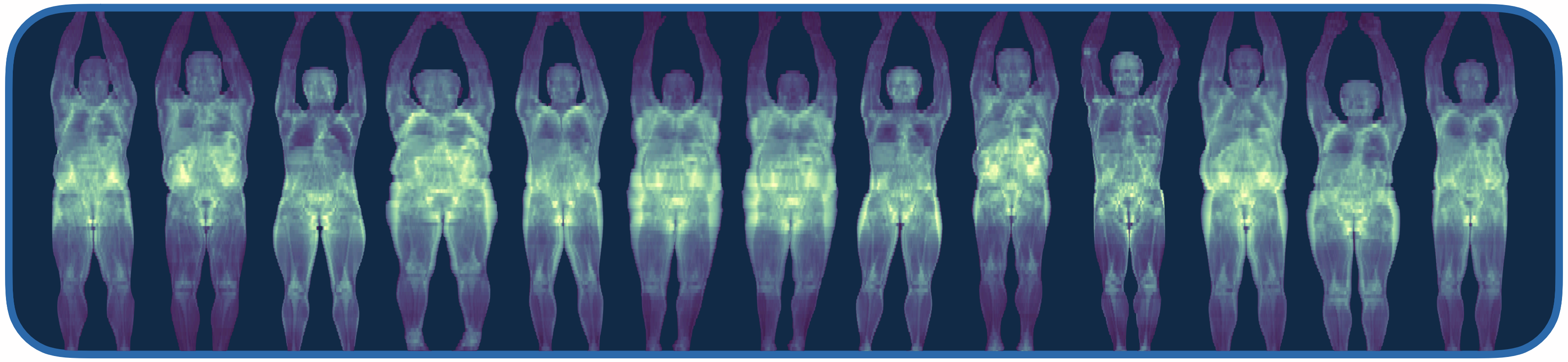
.... → ....

Transforming medical data into valuable insights is a key challenge for health care

1 Korolev et al. 2017 2 Ardila et al. 2019 3 Gulshan et al. 2016 4 Rajpurkar et al. 2017

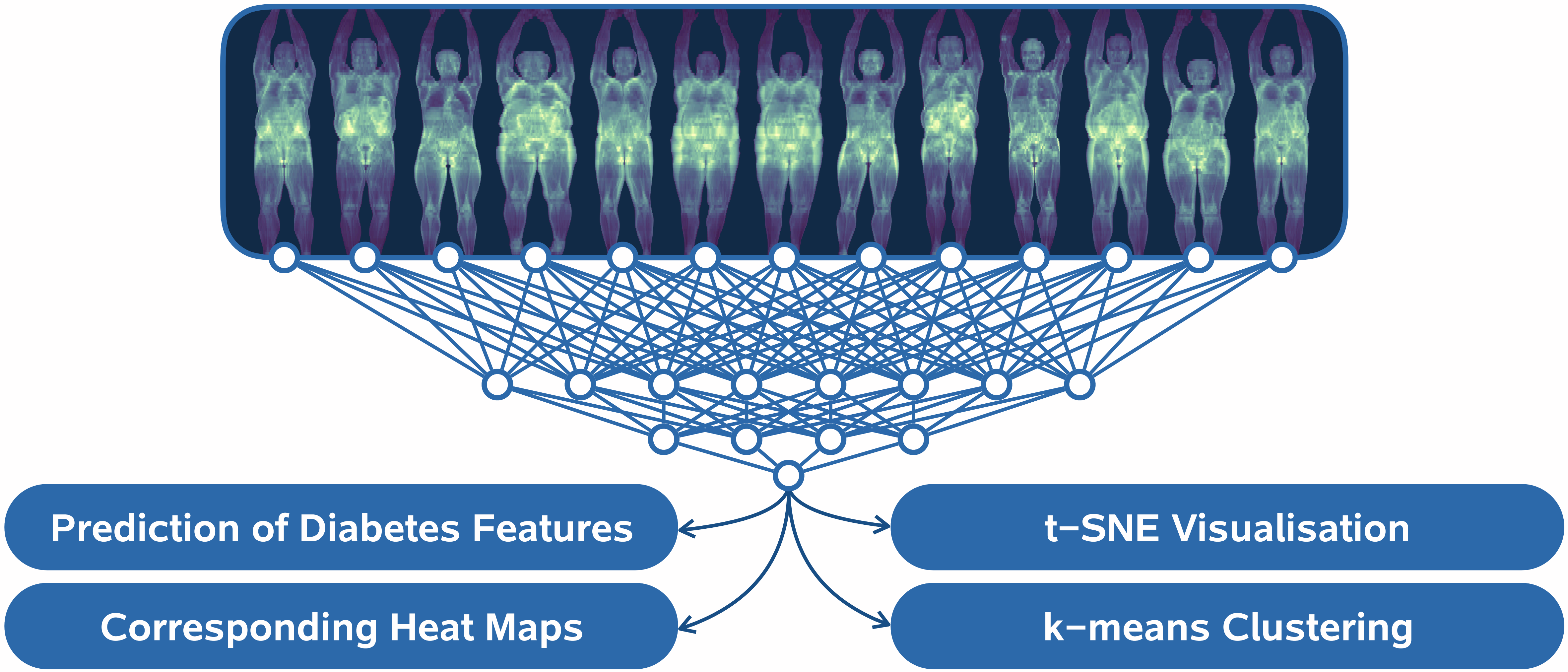
Deep Learning Diabetes Analysis

Full−body MRT dataset provides unique opportunity to utilise Deep Learning



Deep Learning Diabetes Analysis

Full–body MRT dataset provides unique opportunity to utilise Deep Learning



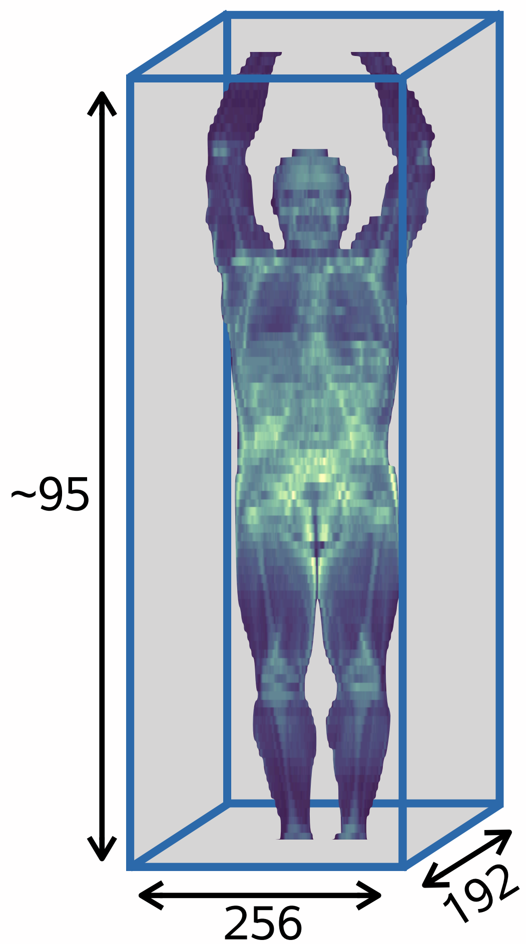
Target–specific heat maps are generated to study potential correlations of certain body regions to specific Diabetes related features

Data & Preprocessing

Preprocessing of MRT Scans

Augmentation of bodies ensures location−independent feature detection

Interpolation

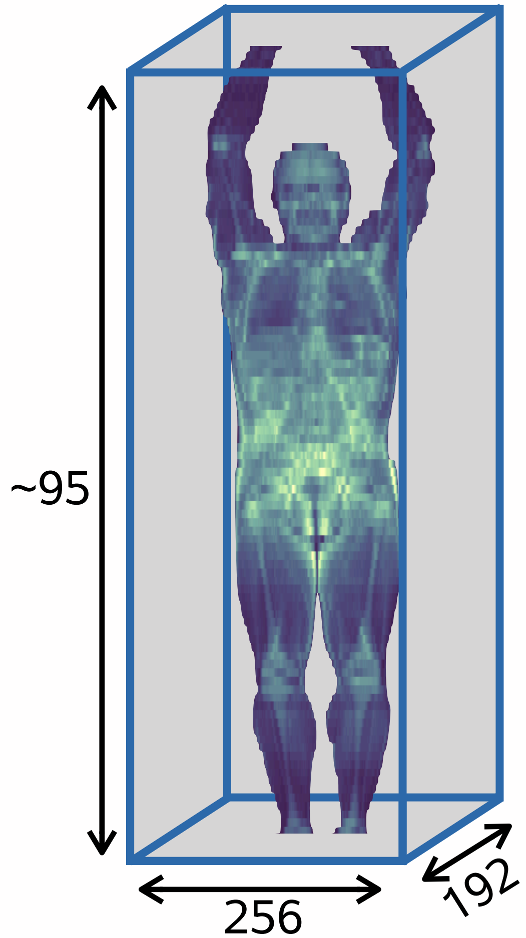


Bodies interpolated and down−sampled
to same dimensions: [85,110,135]

Preprocessing of MRT Scans

Augmentation of bodies ensures location−independent feature detection

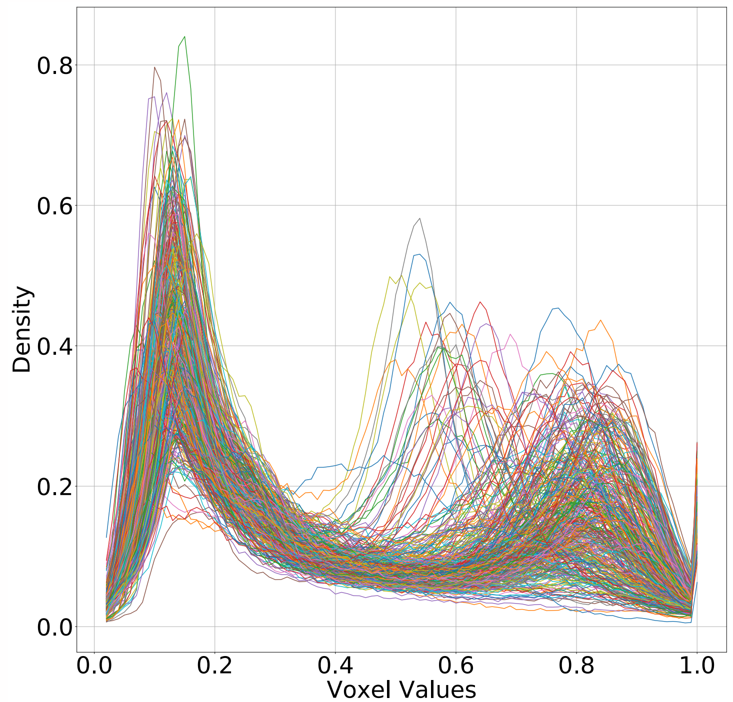
Interpolation



Bodies interpolated and down−sampled to same dimensions: [85,110,135]

Scaling/ Normalisation

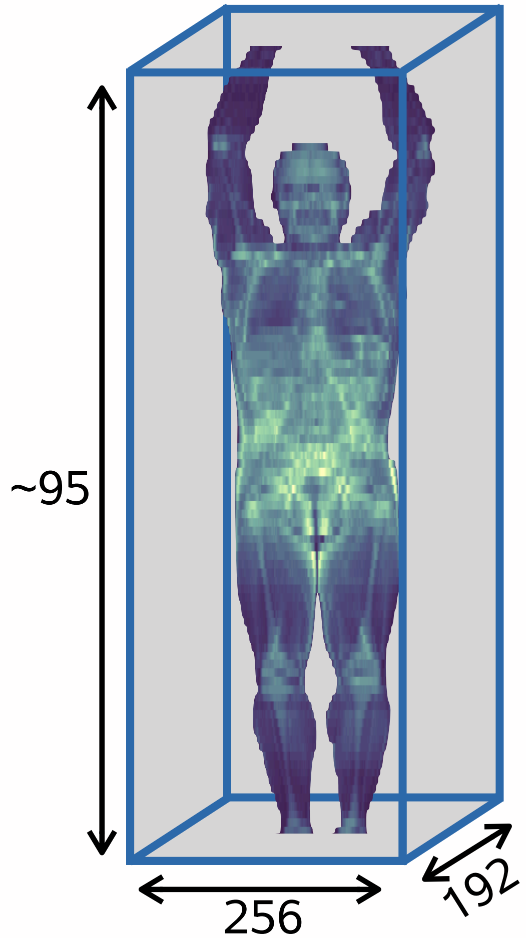
- Shifted / scaled to mean=0, var=1
- Shifted to positive values (0−padding around bodies)
- Normalised voxel distribution:



Preprocessing of MRT Scans

Augmentation of bodies ensures location−independent feature detection

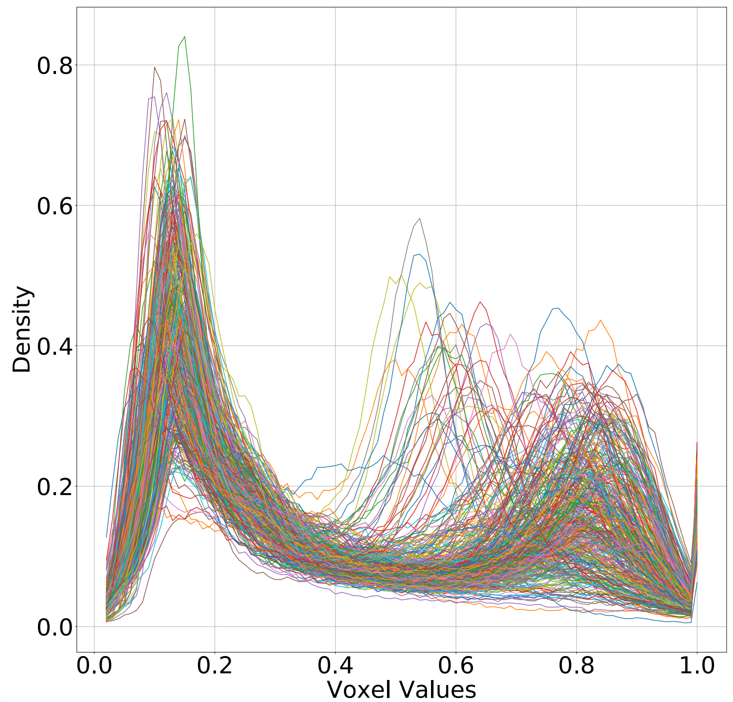
Interpolation



Bodies interpolated and down−sampled to same dimensions: [85,110,135]

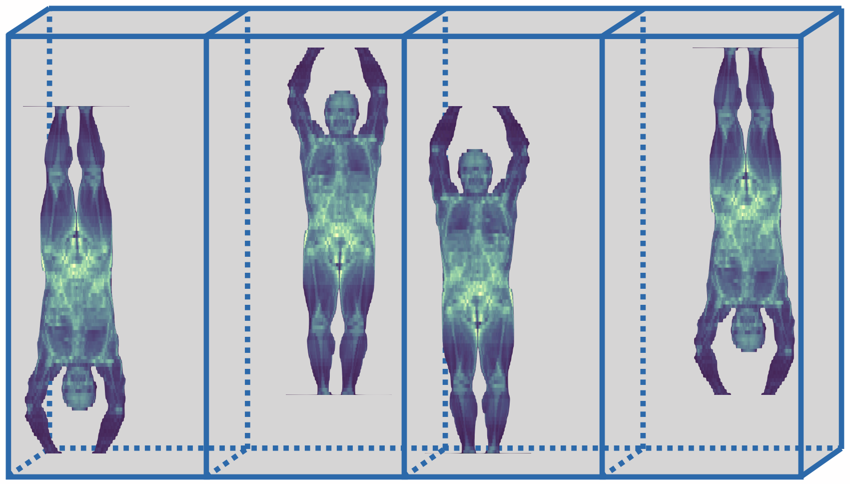
Scaling/ Normalisation

- Shifted / scaled to mean=0, var=1
- Shifted to positive values (0−padding around bodies)
- Normalised voxel distribution:



Augmentation

- Gaussian noise on non−zero values
- Random flipping (all axes)
- Random additional padding in all dimensions



Data Pipeline

Three stratified data folds monitor generalisation capabilities and provide final score

Target labels

- Diabetes (D_α)
- Prediabetes (D_β)
- Diabetes IFG+IGT (D_γ)
- Gender
- Age
- Body Mass Index
- Insulin Sensitivity
- HbA1c

Data Pipeline

Three stratified data folds monitor generalisation capabilities and provide final score

Target labels

- Diabetes (D_α)
- Prediabetes (D_β)
- Diabetes IFG+IGT (D_γ)
- Gender
- Age
- Body Mass Index
- Insulin Sensitivity
- HbA1c

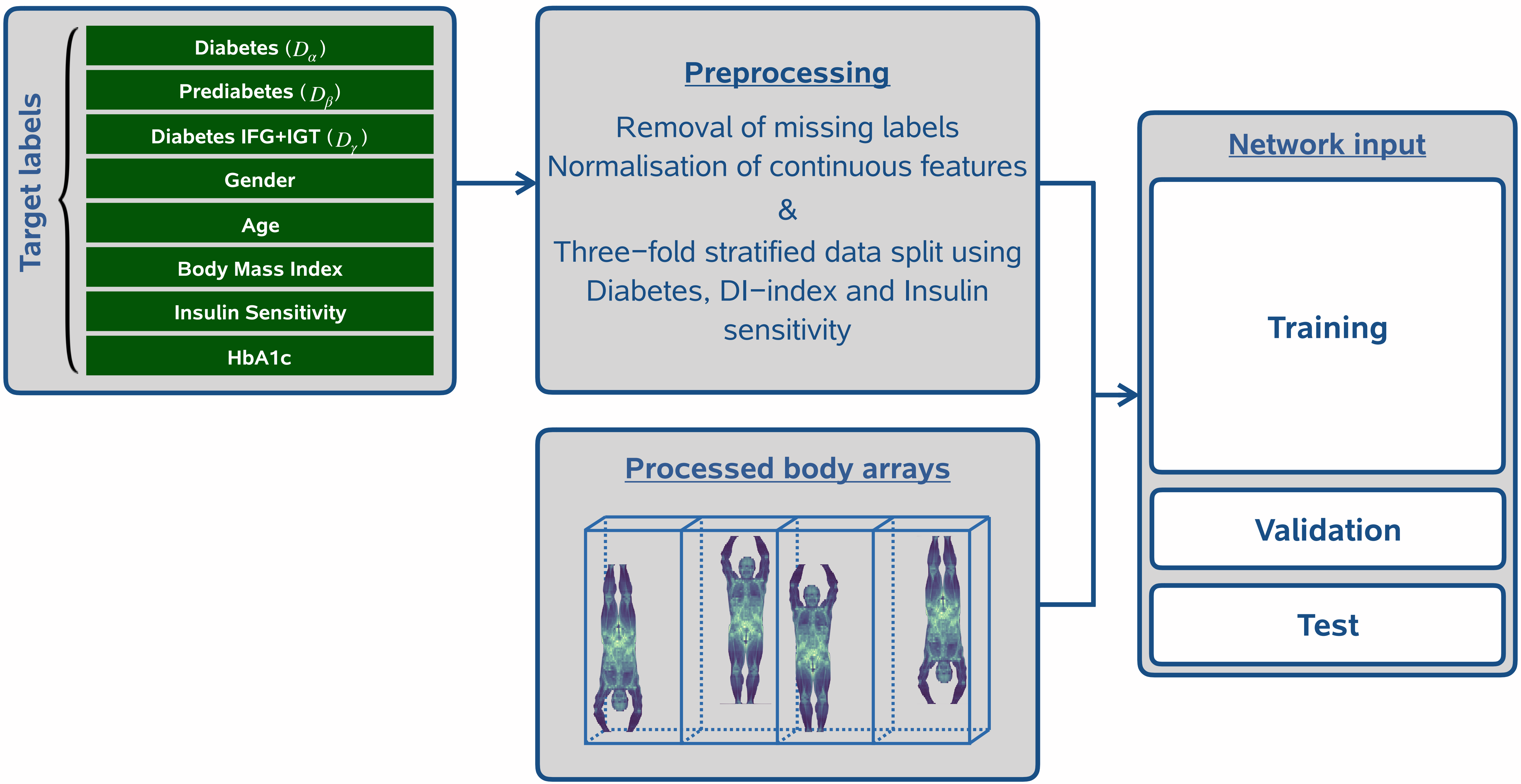
→

Preprocessing

Removal of missing labels
Normalisation of continuous features
&
Three−fold stratified data split using Diabetes, DI−index and Insulin sensitivity

Data Pipeline

Three stratified data folds monitor generalisation capabilities and provide final score

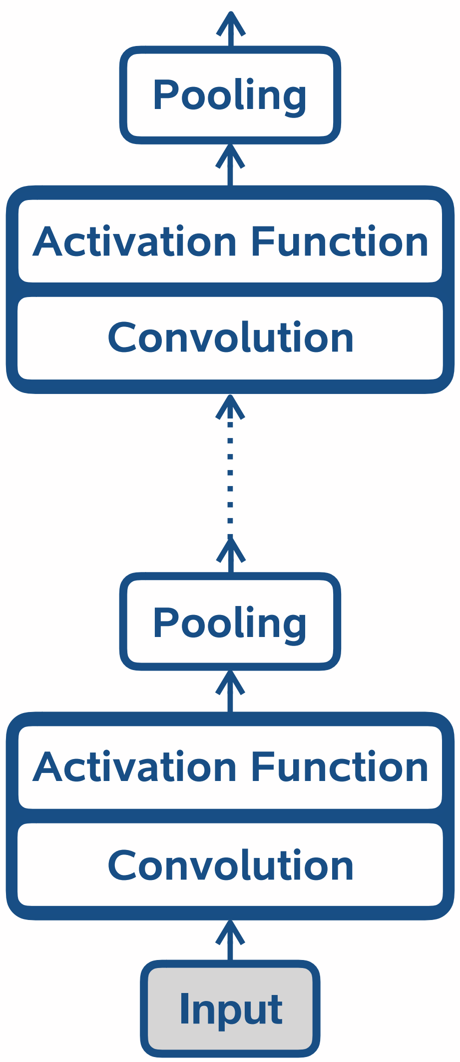


Model Architecture

Densely Connected Convolutional Networks

Deep and computationally efficient neural network for image recognition tasks

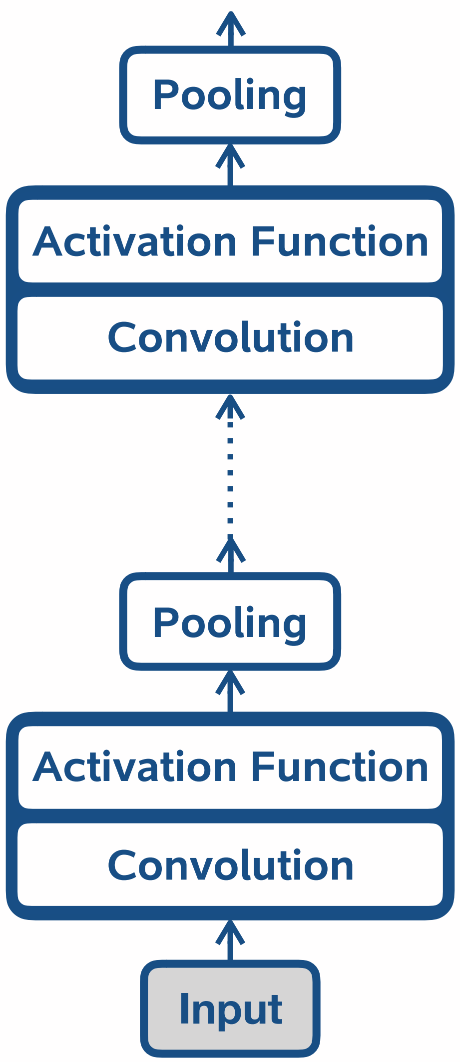
Standard Approach:
Convolutional Neural Network



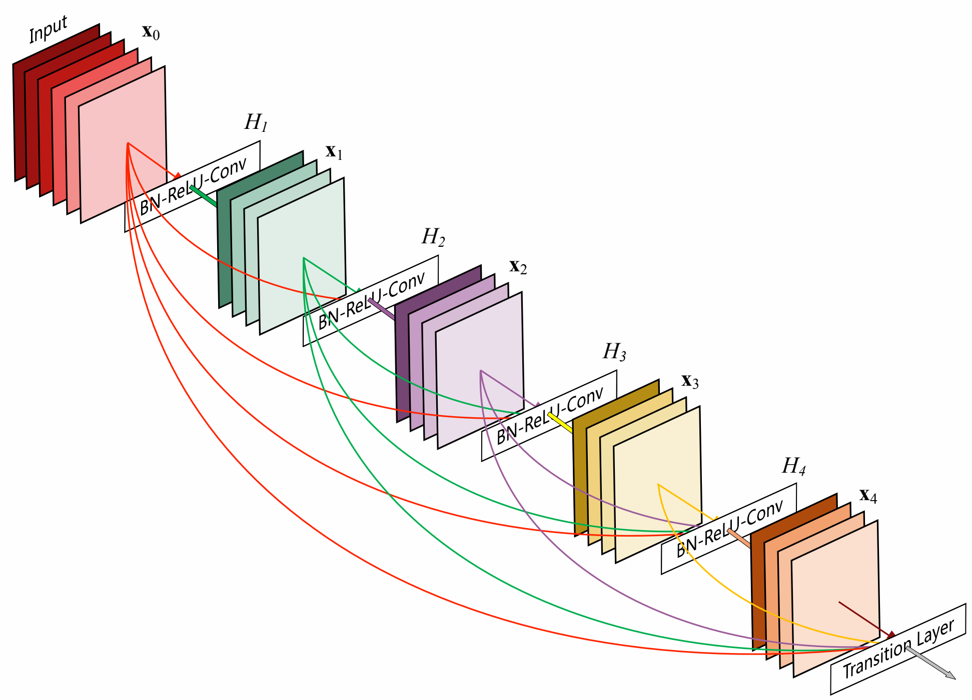
Densely Connected Convolutional Networks

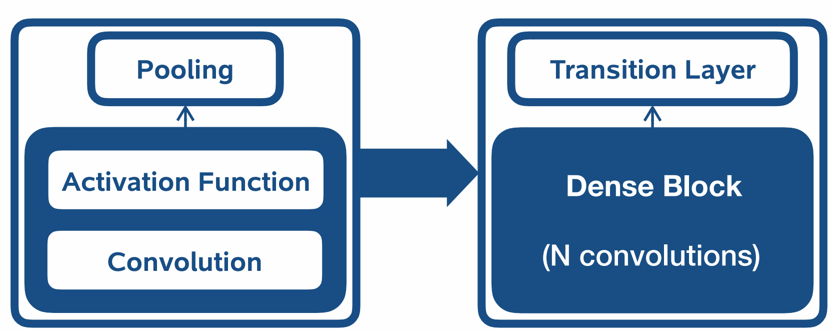
Deep and computationally efficient neural network for image recognition tasks

Standard Approach:
Convolutional Neural Network



Huang et al. (2017):
DenseNet

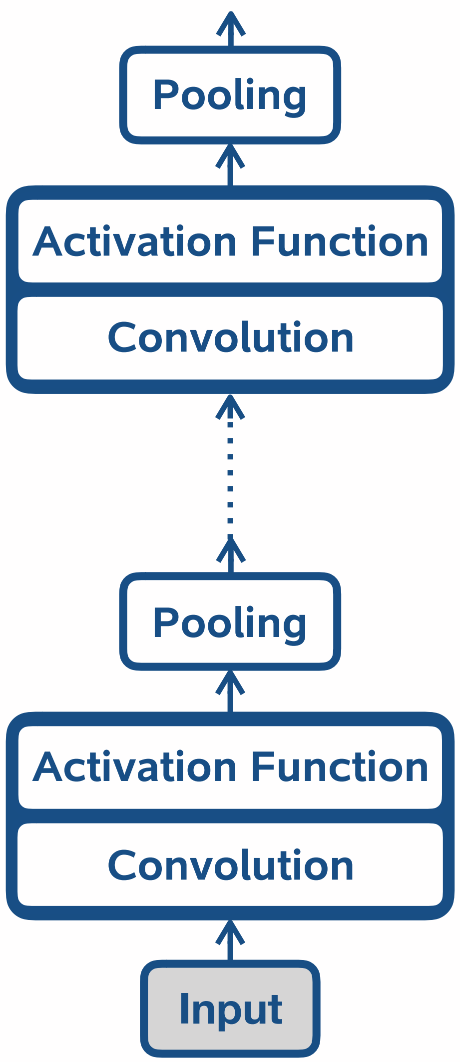




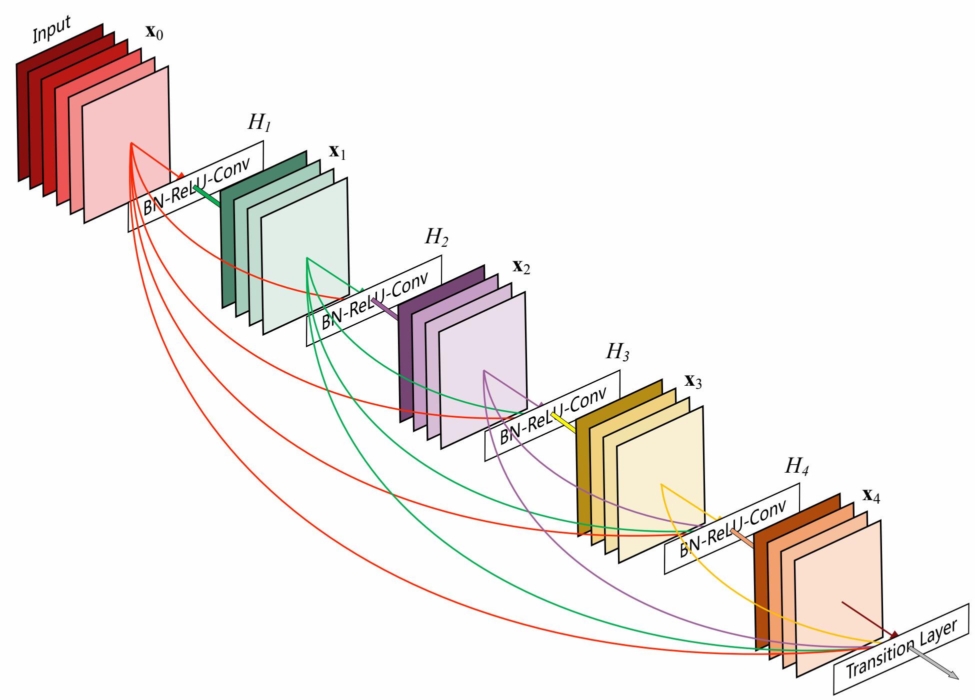
Densely Connected Convolutional Networks

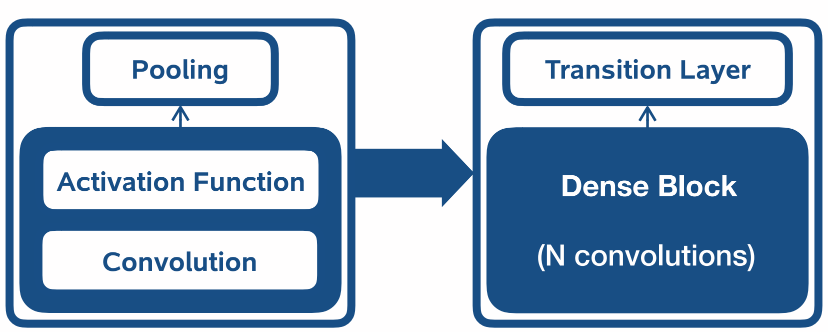
Deep and computationally efficient neural network for image recognition tasks

Standard Approach:
Convolutional Neural Network



Huang et al. (2017):
DenseNet





Advantages

- **Improved flow of information**
- **Very deep models possible**
- **Feature Reuse**
- **Efficient training**
- **Bottleneck layers**
- **Mitigates overfitting**
- **Explicit compression possible**

Model Architecture

From MRT scans to all outputs

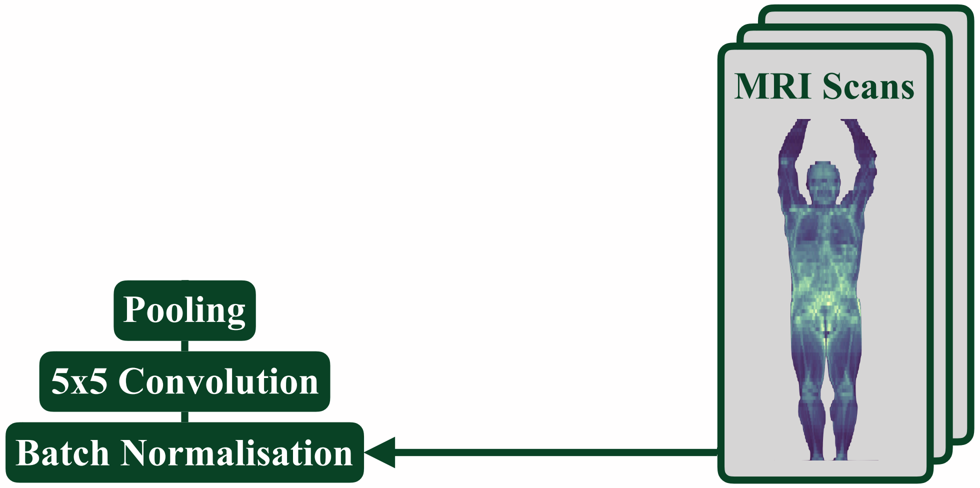
MRT scan inputs [85,110,135]

Model Architecture

From MRT scans to all outputs

MRT scan inputs [85,110,135]

Initial convolution and pooling



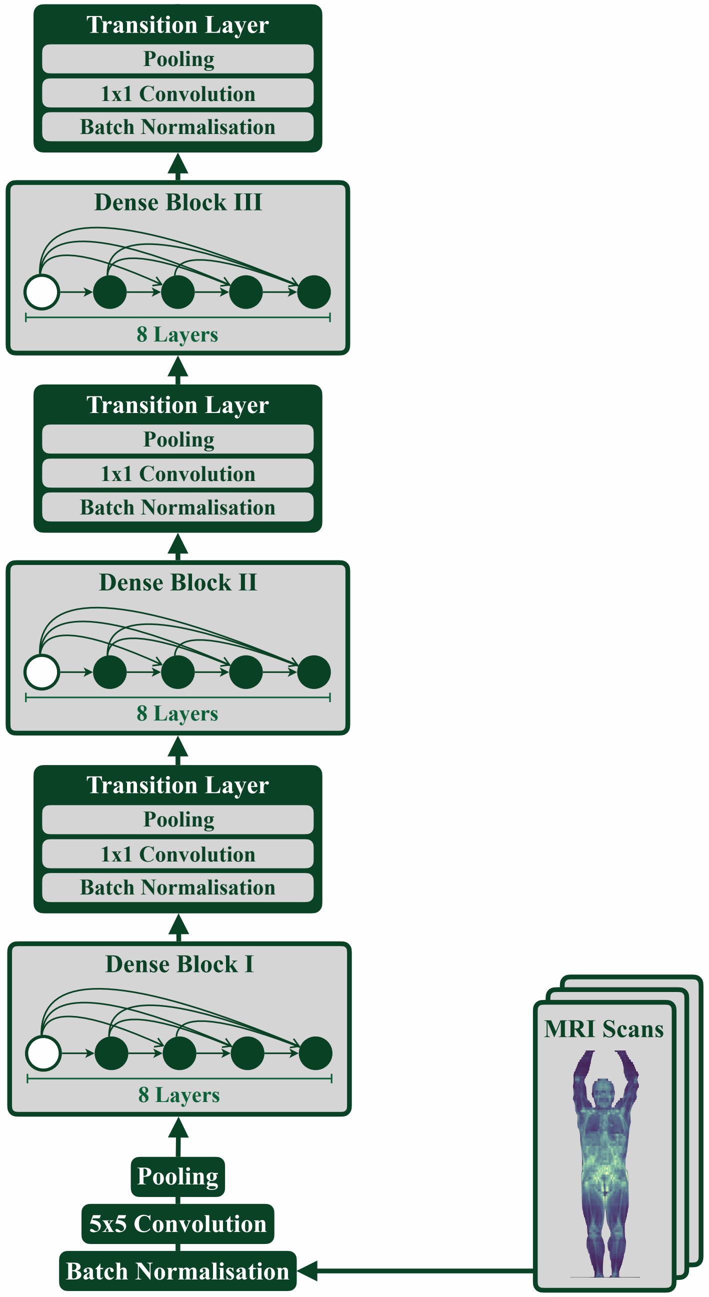
Model Architecture

From MRT scans to all outputs

MRT scan inputs [85,110,135]

Initial convolution and pooling

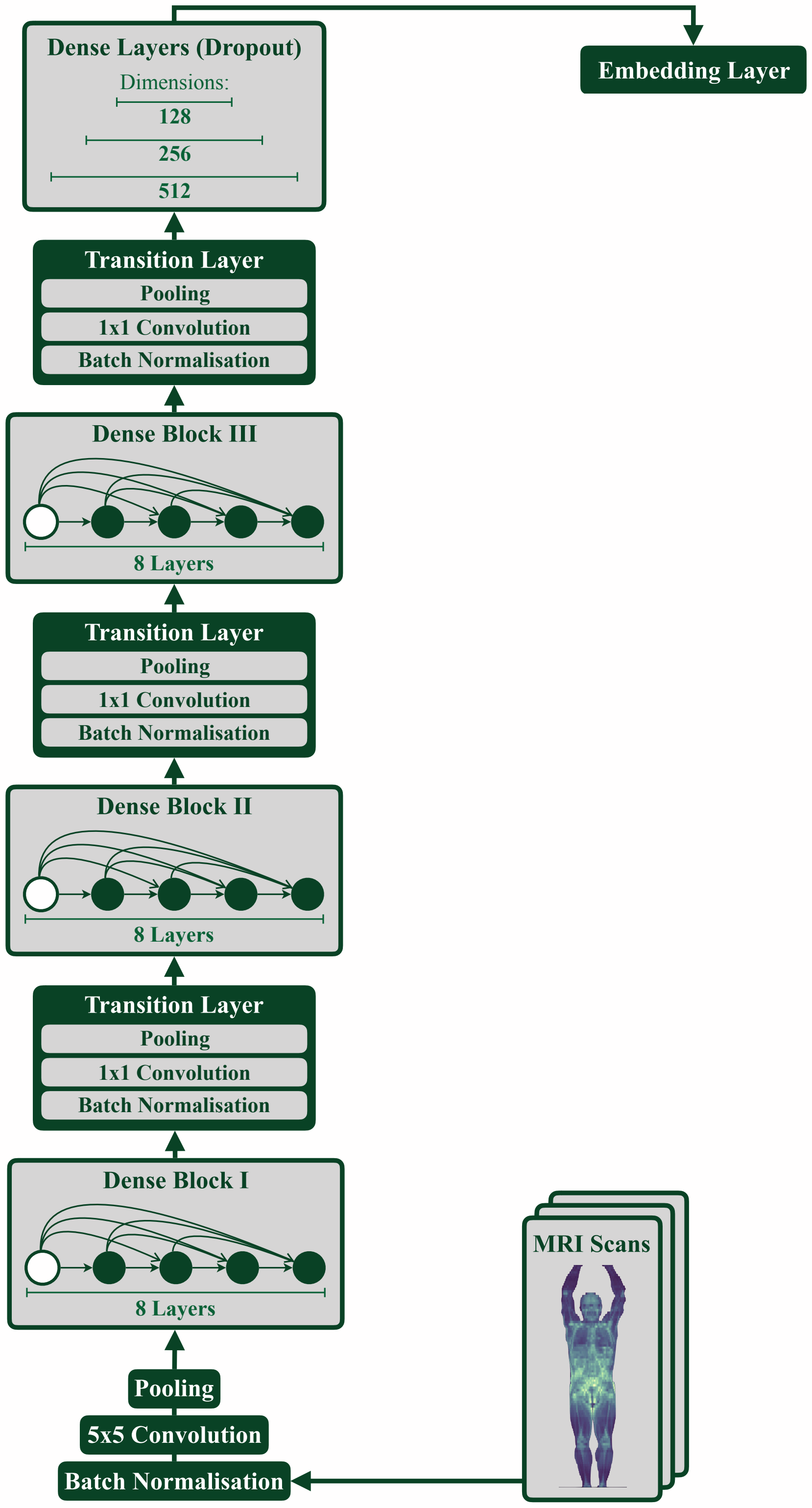
DenseNet feature extraction



Model Architecture

From MRT scans to all outputs

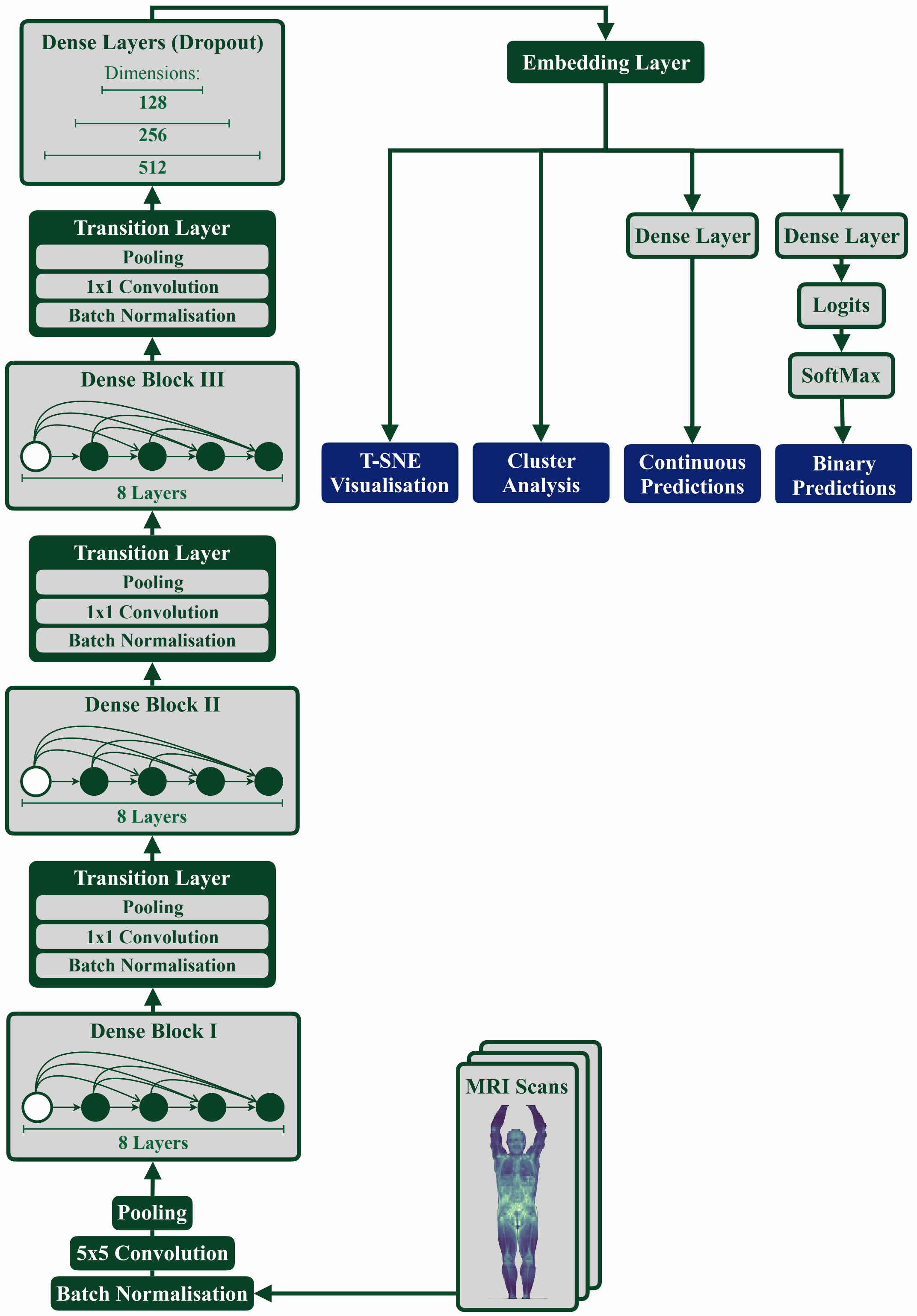
- MRT scan inputs [85,110,135]
- Initial convolution and pooling
- DenseNet feature extraction
- **Mapping to embedding layer [128]**



Model Architecture

From MRT scans to all outputs

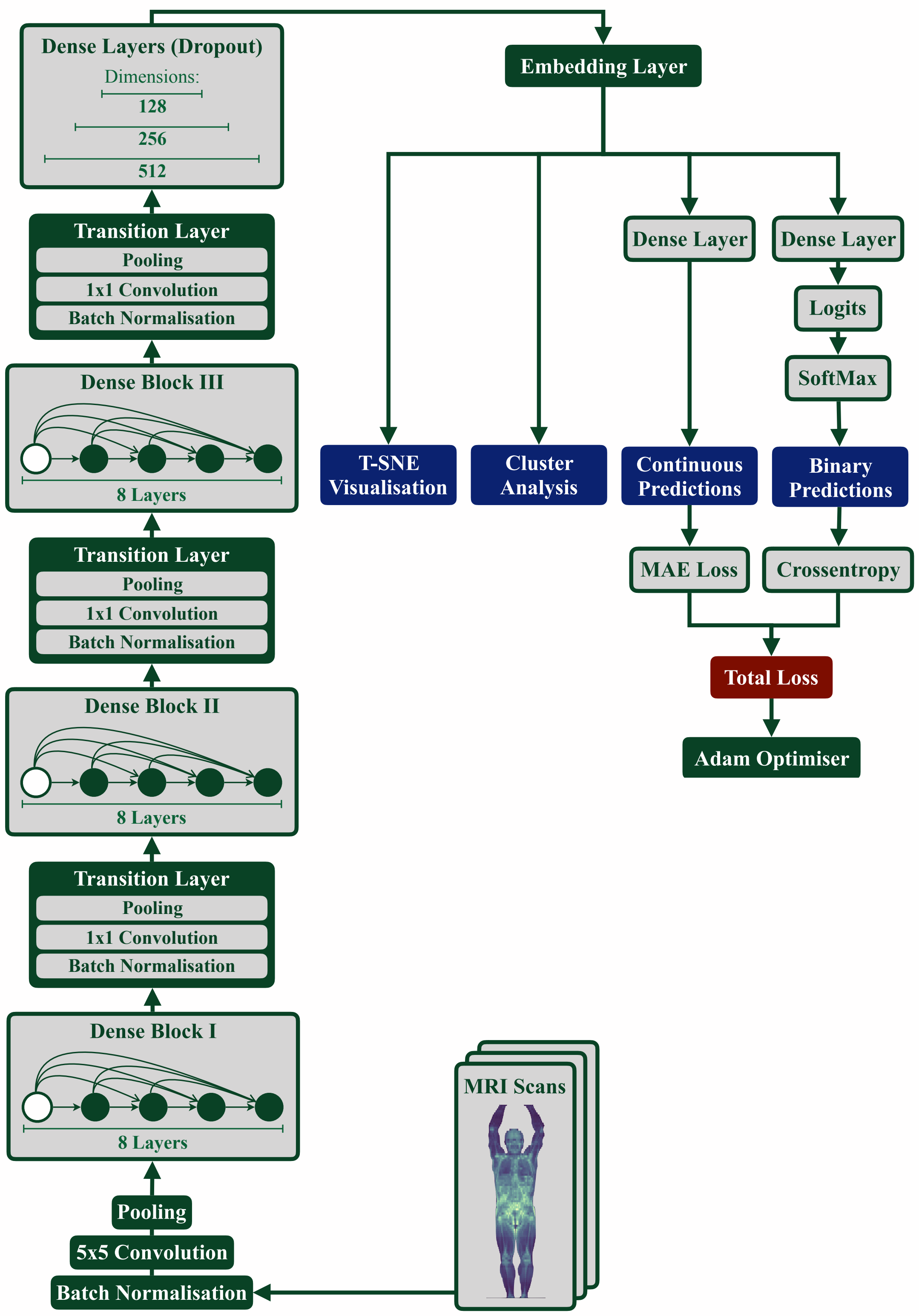
- MRT scan inputs [85,110,135]
- Initial convolution and pooling
- DenseNet feature extraction
- Mapping to embedding layer [128]
  - ➡ Input to all subsequent steps
    - ➡ Predictions | t–SNE | k–means



Model Architecture

From MRT scans to all outputs

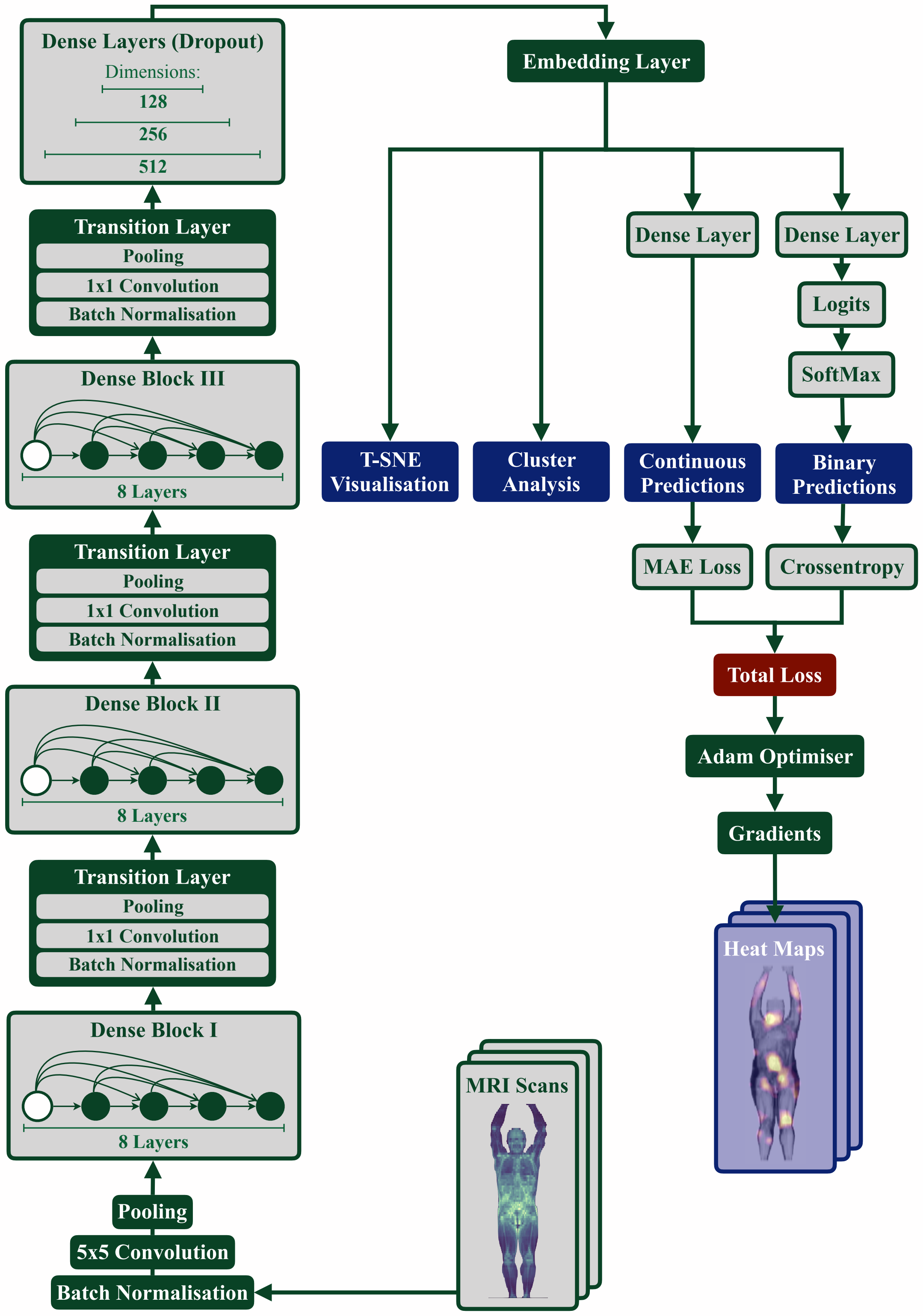
- MRT scan inputs [85,110,135]
- Initial convolution and pooling
- DenseNet feature extraction
- Mapping to embedding layer [128]
  - ➡ Input to all subsequent steps
    - ➡ Predictions | t–SNE | k–means
- **Crossentropy– and MAE–loss**
- **Adam optimiser**



Model Architecture

From MRT scans to all outputs

- MRT scan inputs [85,110,135]
- Initial convolution and pooling
- DenseNet feature extraction
- Mapping to embedding layer [128]
  - ➡ Input to all subsequent steps
    - ➡ Predictions | t−SNE | k−means
- Crossentropy− and MAE−loss
- Adam optimiser
- **Gradient maps**

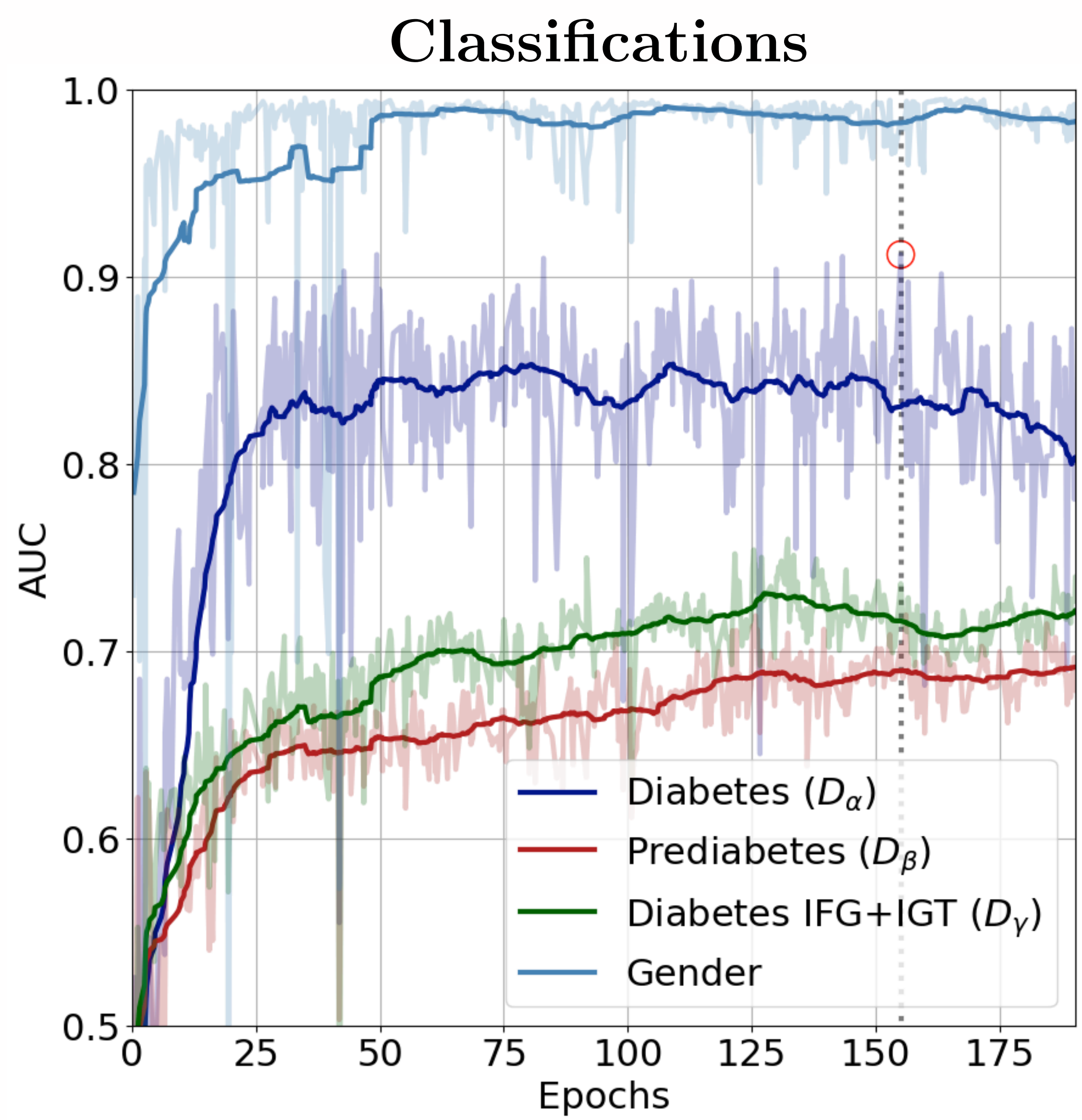


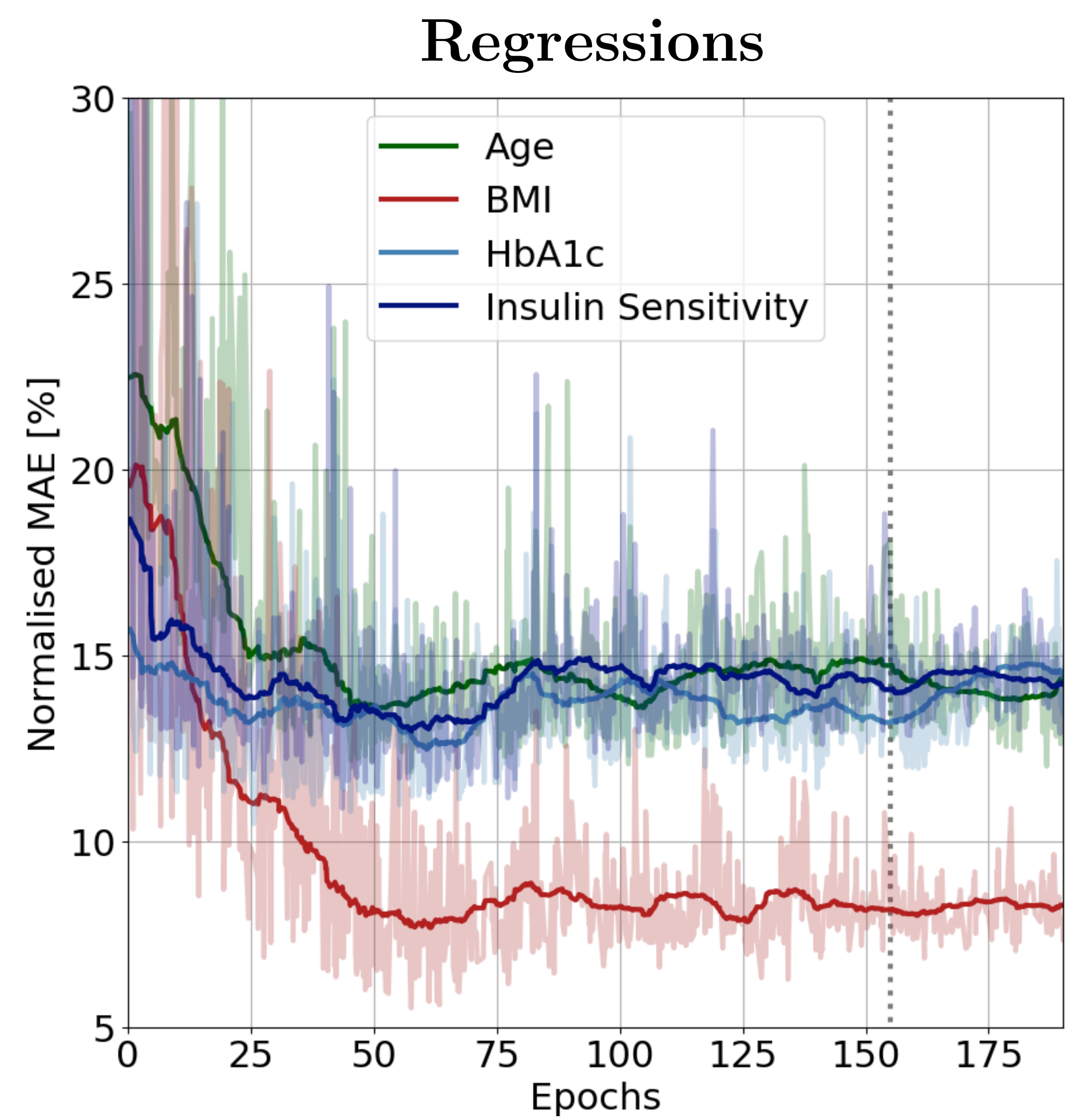
Results

Training

Performance on the Training Set

Model selected according to highest diabetes AUC





HbA1c and Insulin sensitivity prove to be challenging

Benedikt Dietz
August 9th, 2019

Predictive Performance

Performance Summary

Regression

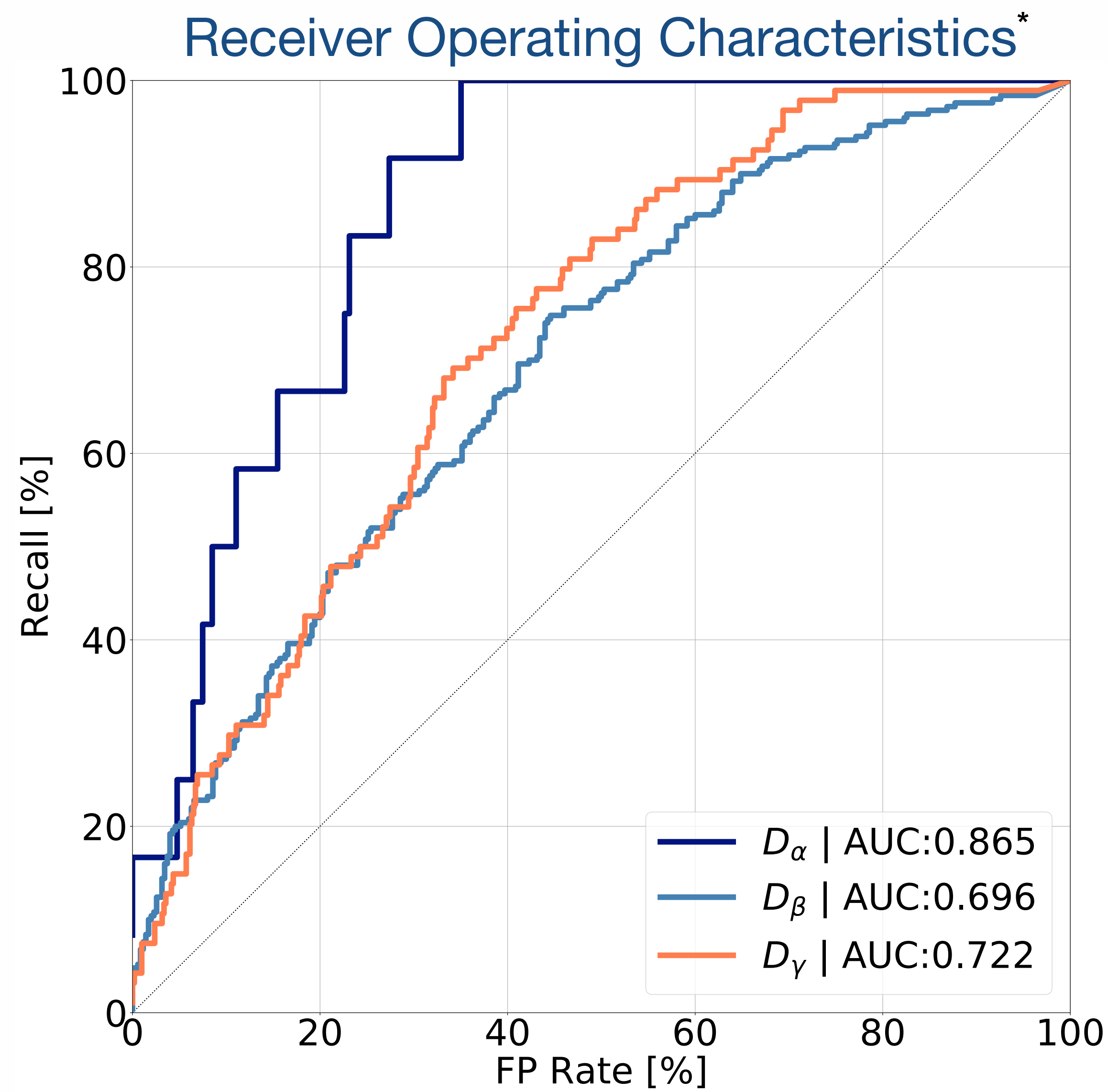
- Body Mass Index performs best
- Insulin Sensitivity and HbA1c are challenging

	MAE (norm.)	MAE (total)
Age	2%	
Body Mass Index	15%	
HbA1c	45%	
Insulin Sensitivity		

Classification

- AUROC used as key metric
- Gender AUROC converges to >99%
- Diabetes AUROC varies according to prevalence

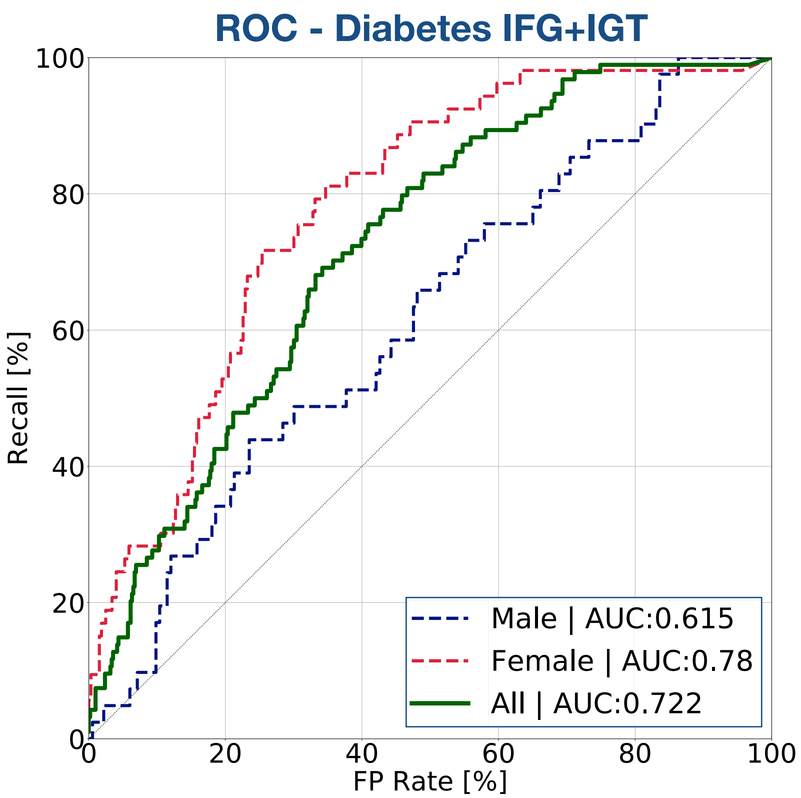
	Prevalence	AUROC
Diabetes* D_α	2%	0.87
Diabetes IFG+IGT D_γ	15%	0.72
Prediabetes D_β	45%	0.70
Gender		0.99

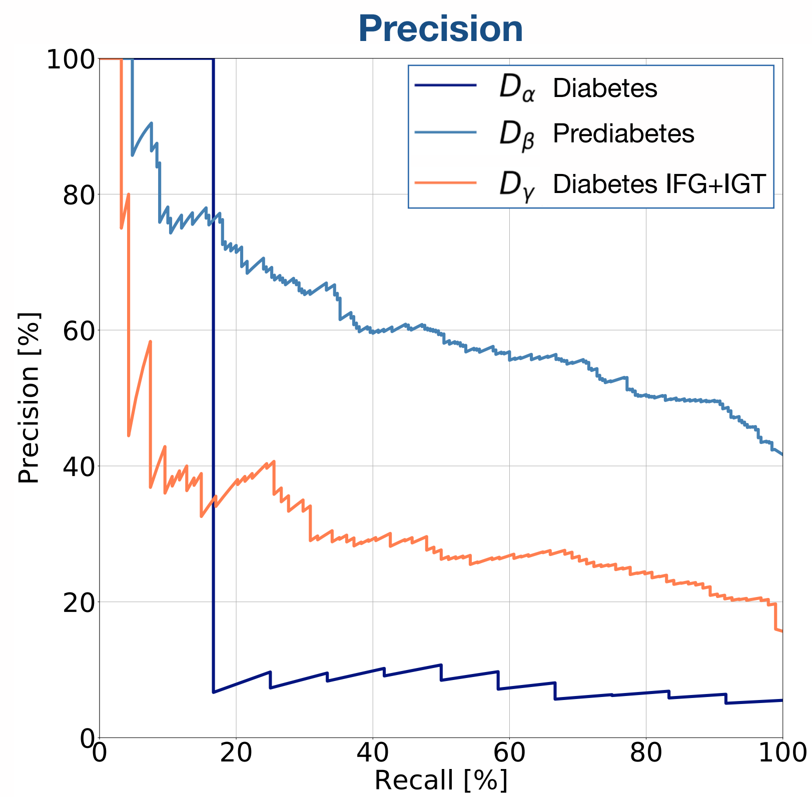


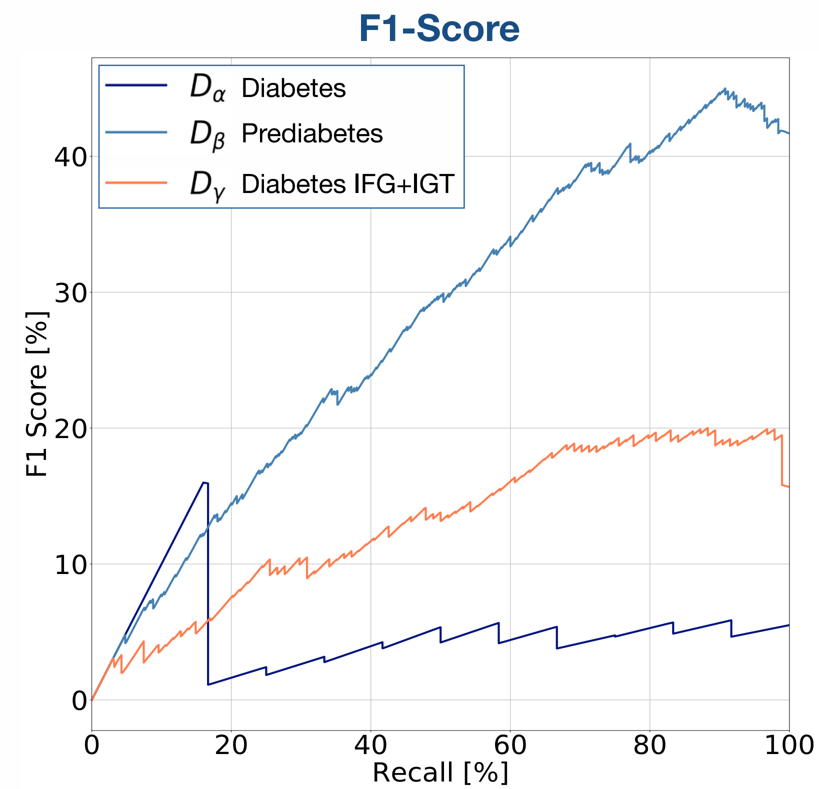
*Computed on test– and validation set due to critically low number of positives

Diabetes Classification in Detail

Label imbalance and gender have considerable effects







Female AUROC consistently better

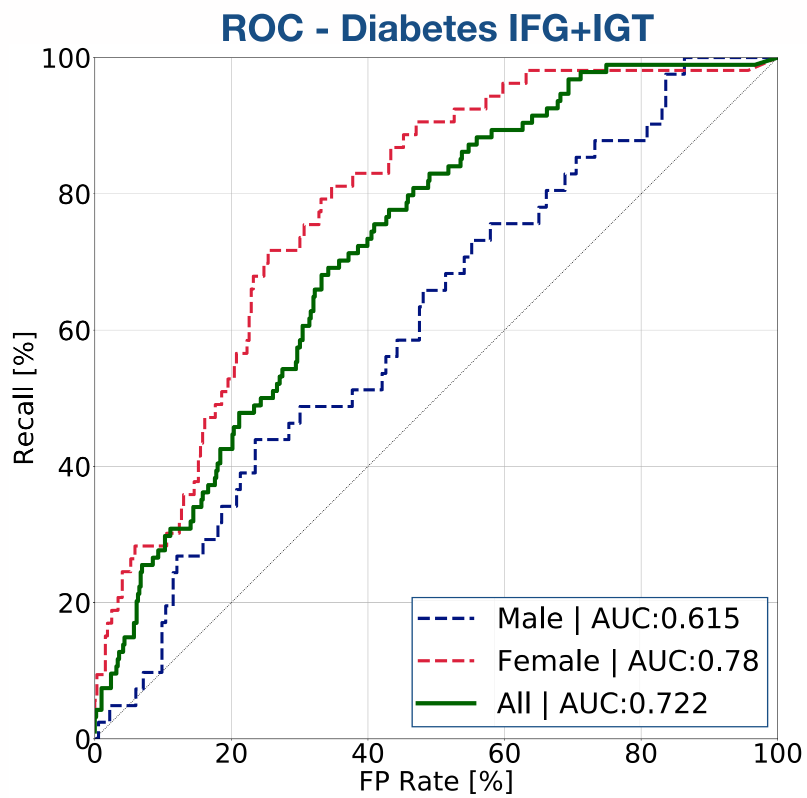
	Male	Female	△
Diabetes	0.80	0.89	**11%**
Diabetes IFG+IGT	0.62	0.78	**27%**
Prediabetes	0.68	0.72	**5%**

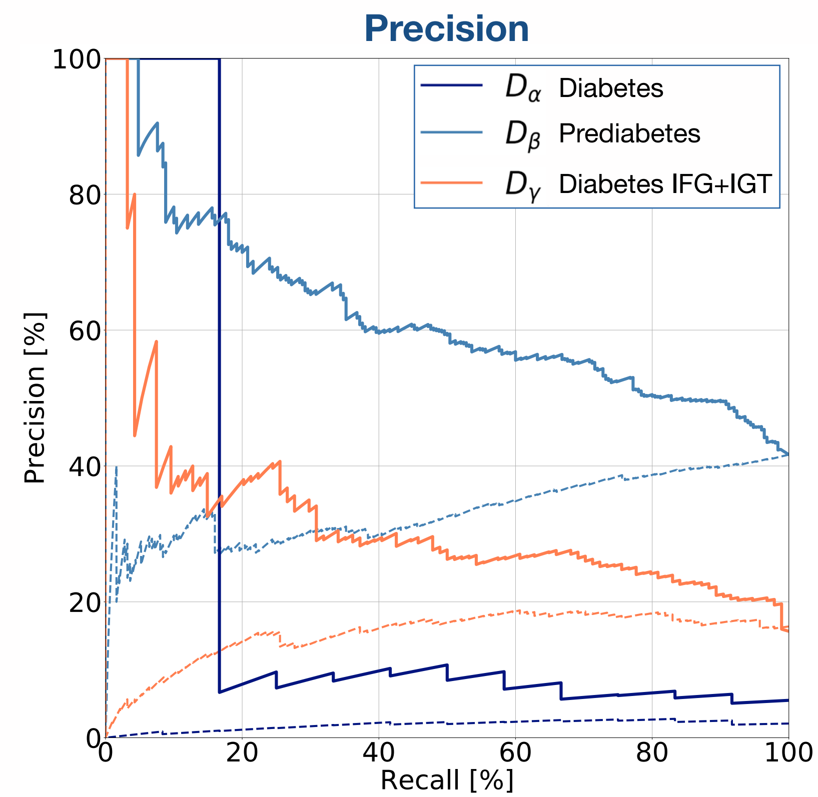
Label imbalance has a harsh effect on precision and F1-score

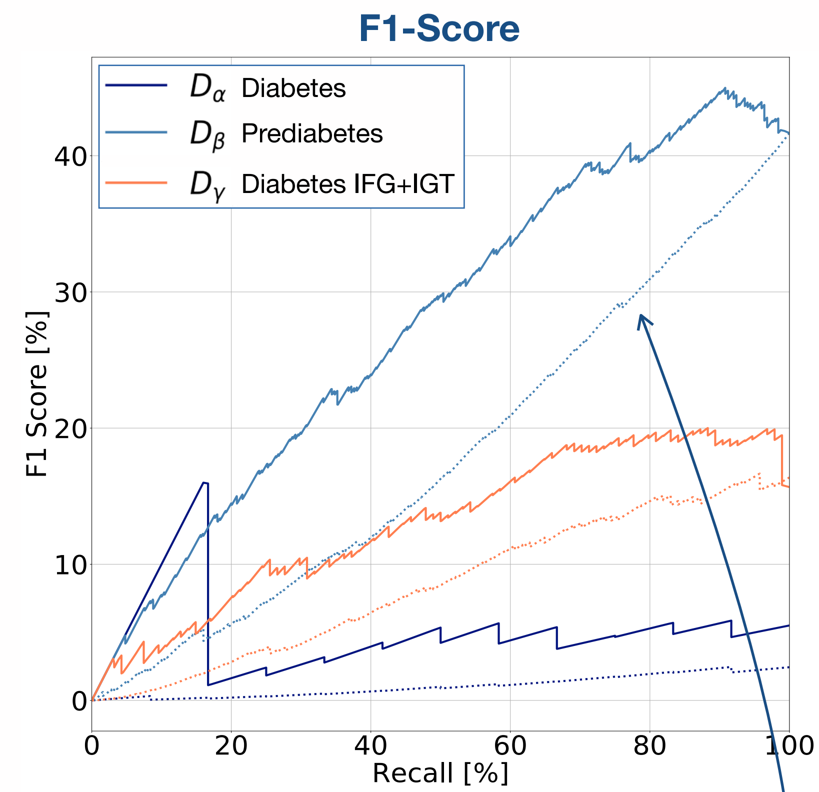
Computed on test– and validation set due to critically low number of positives

Diabetes Classification in Detail

Label imbalance and gender have considerable effects







Female AUROC consistently better

	Male	Female	△
Diabetes	0.80	0.89	**11%**
Diabetes IFG+IGT	0.62	0.78	**27%**
Prediabetes	0.68	0.72	**5%**

Label imbalance has a harsh effect on precision and F1-score

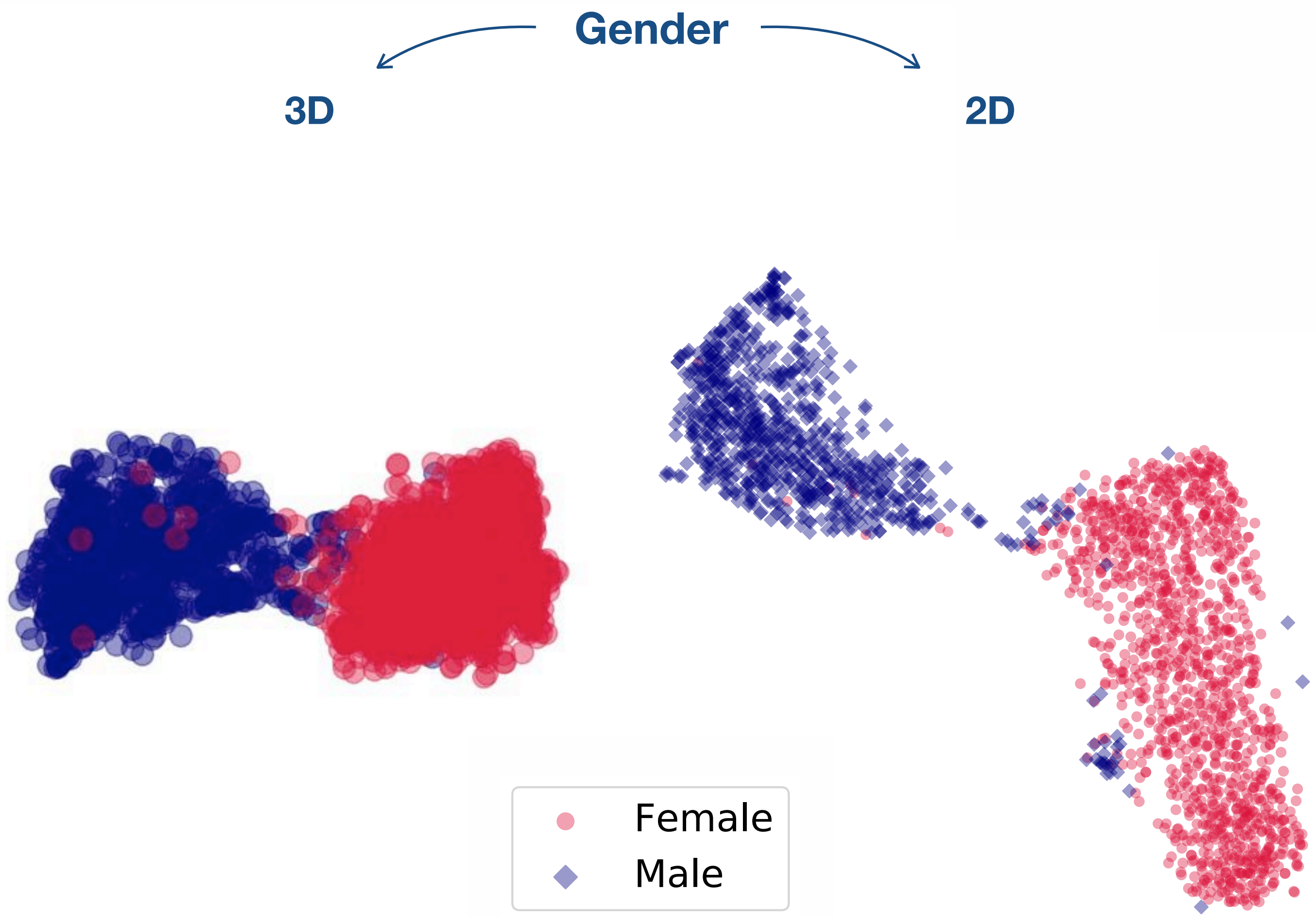
Support Vector Classifier trained on the same training set using BMI and Insulin sensitivity

Computed on test− and validation set due to critically low number of positives

Embedding Space Analysis

t–Distributed Stochastic Neighbour Embedding

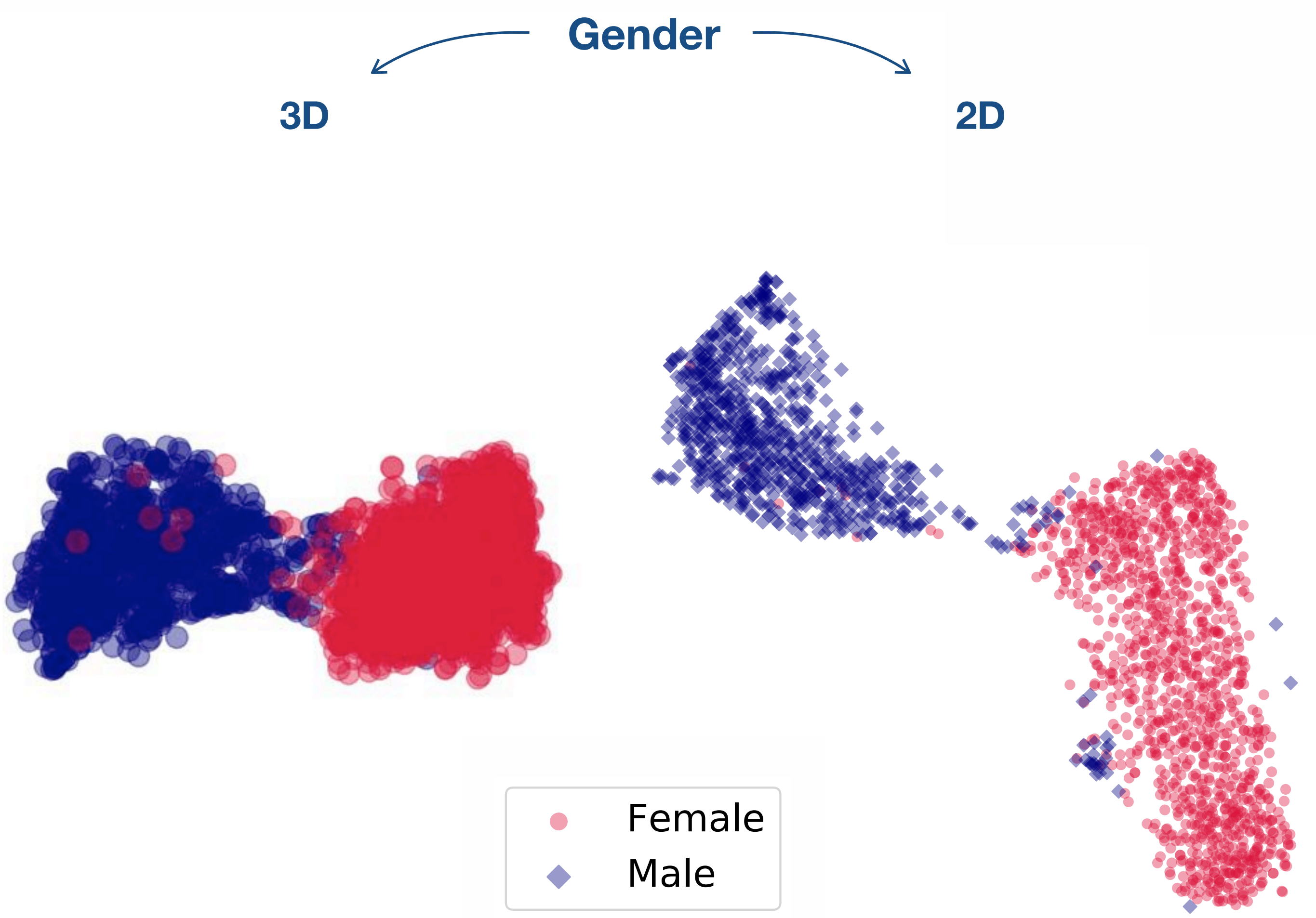
t–SNE maps the high–dimensional representation to 2D for visual analysis



Embedding produces two gender–specific clusters and visualises feature distributions

t–Distributed Stochastic Neighbour Embedding

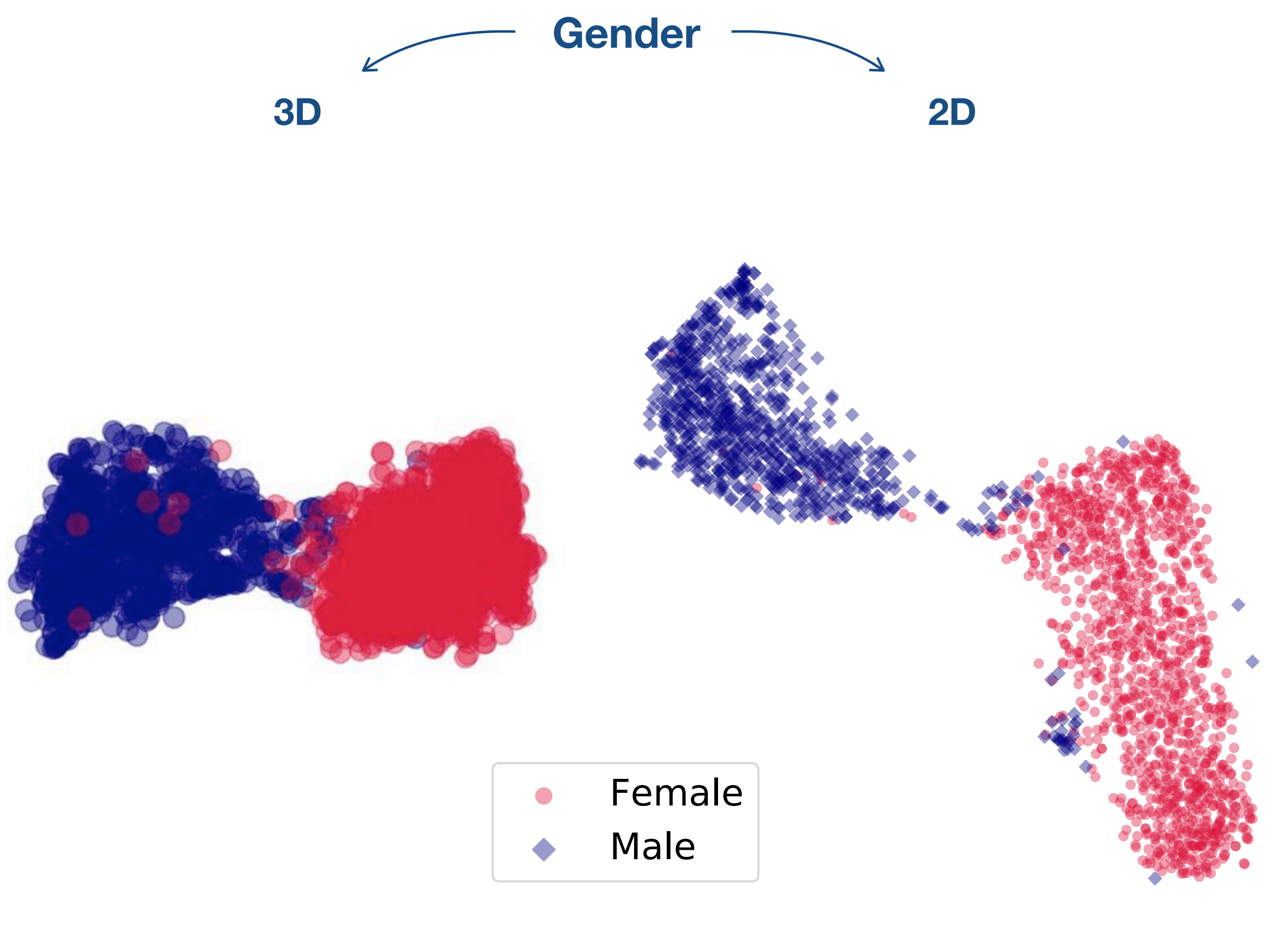
t–SNE maps the high–dimensional representation to 2D for visual analysis

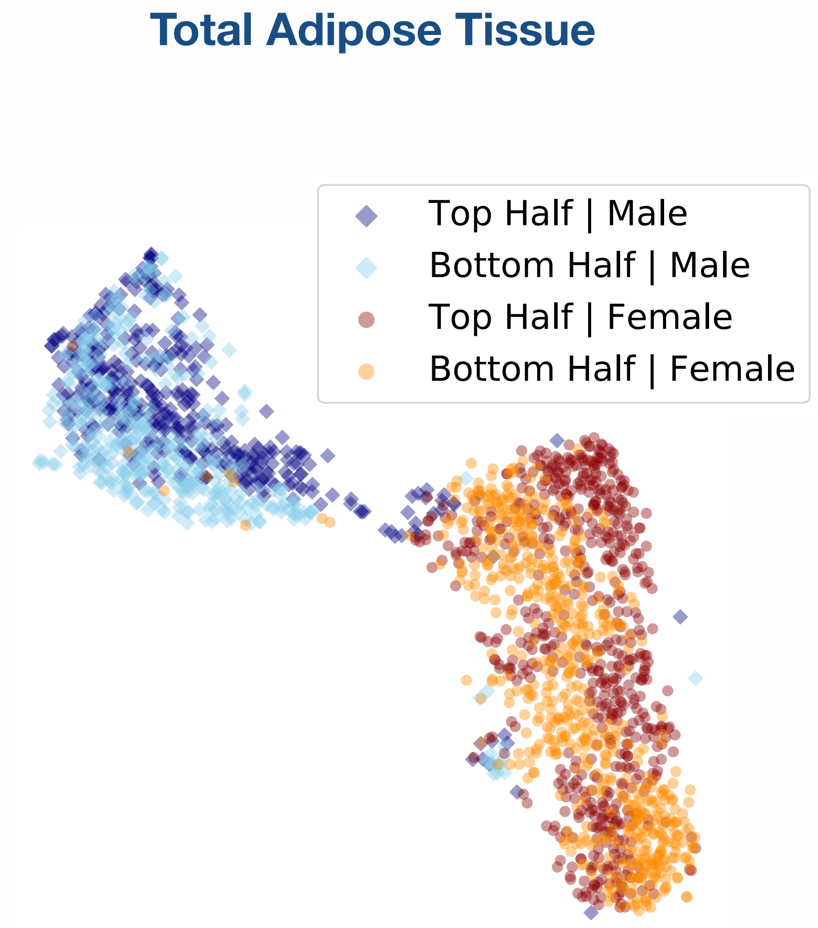


Embedding produces two gender–specific clusters and visualises feature distributions

t–Distributed Stochastic Neighbour Embedding

t–SNE maps the high–dimensional representation to 2D for visual analysis



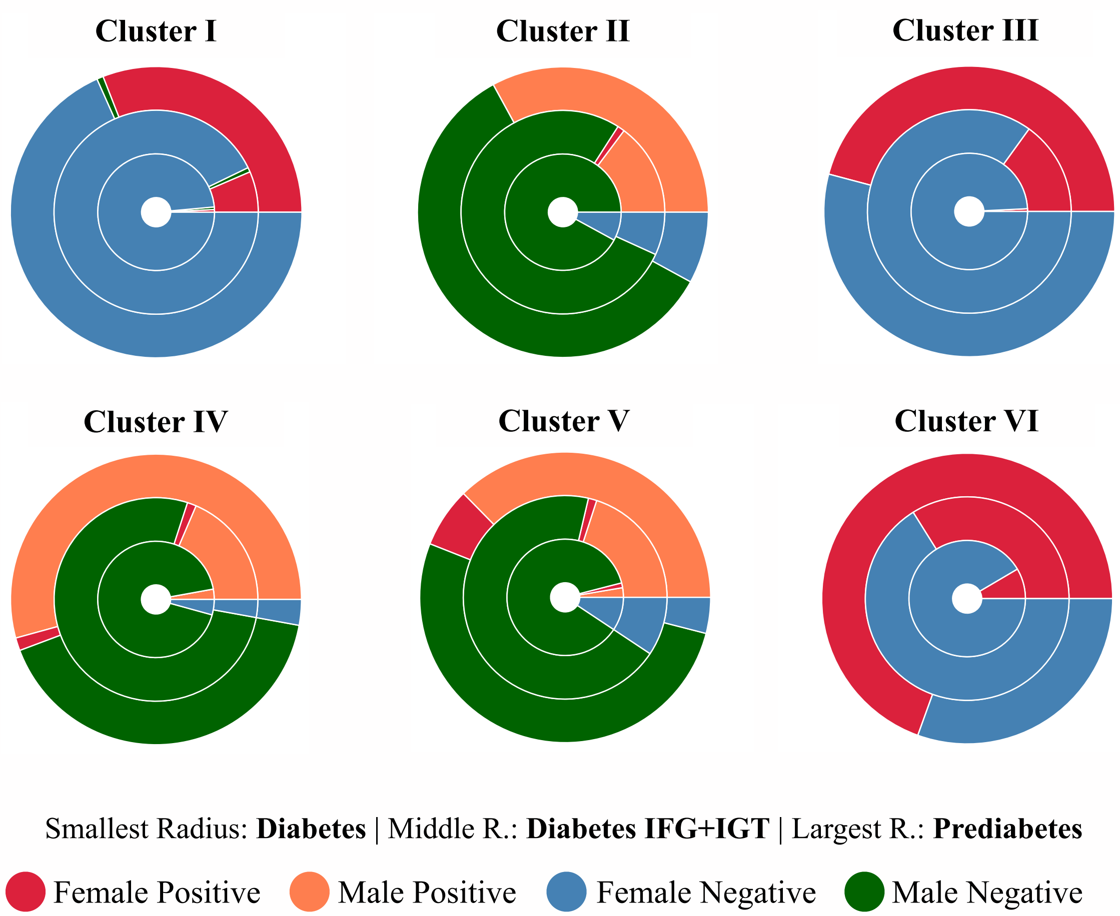


Embedding produces two gender–specific clusters and visualises feature distributions

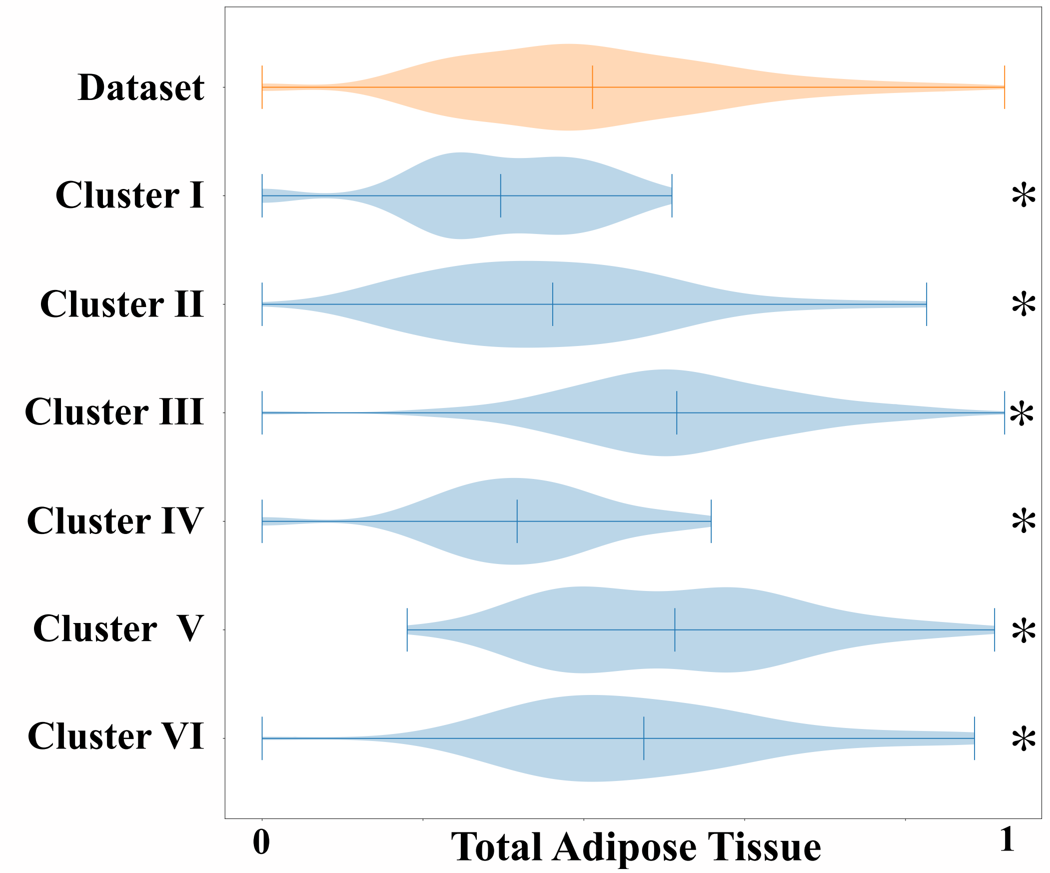
k–means Clustering

Further analysis of the embedding layer

Diabetes Distributions



Continuous Feature Distributions

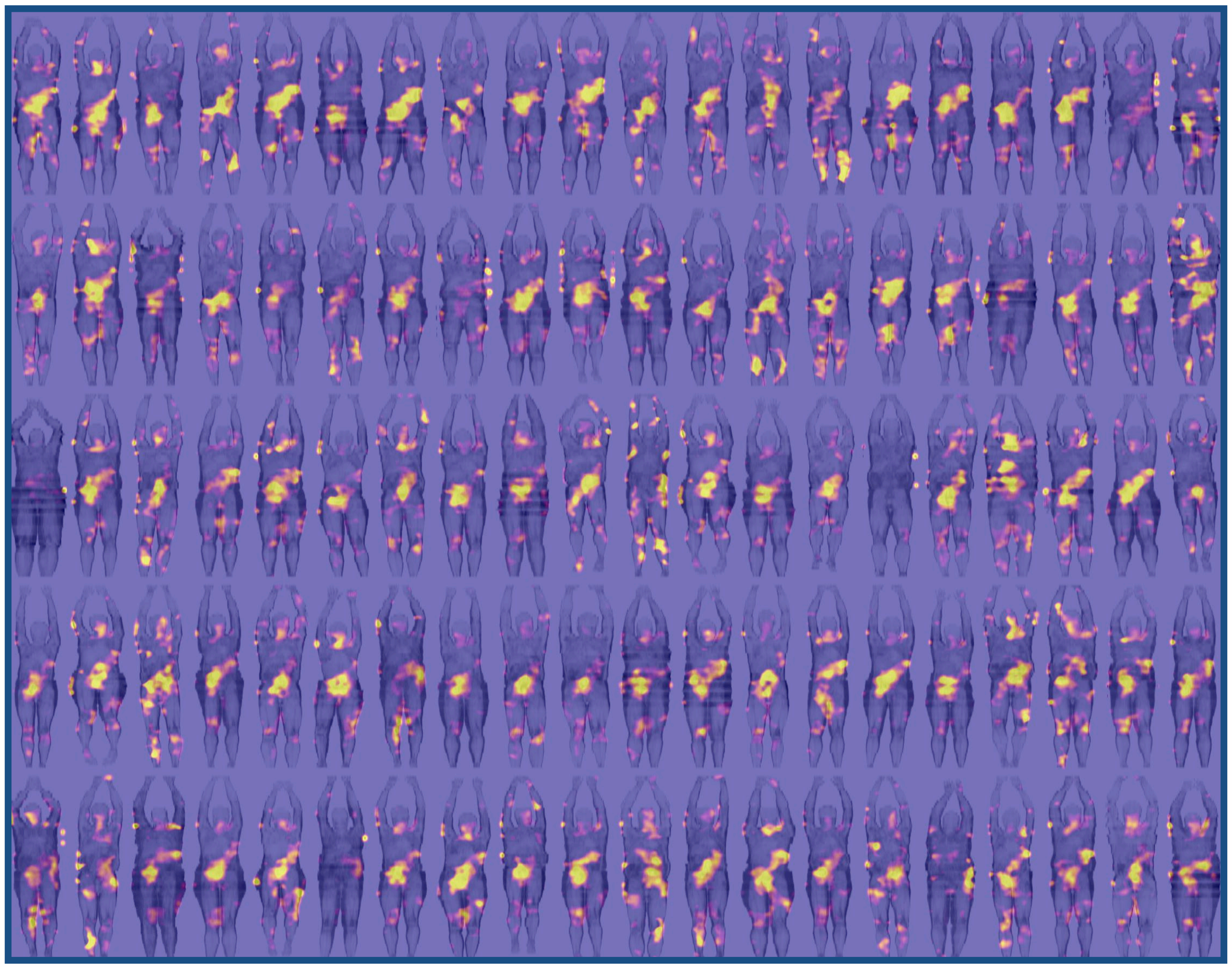


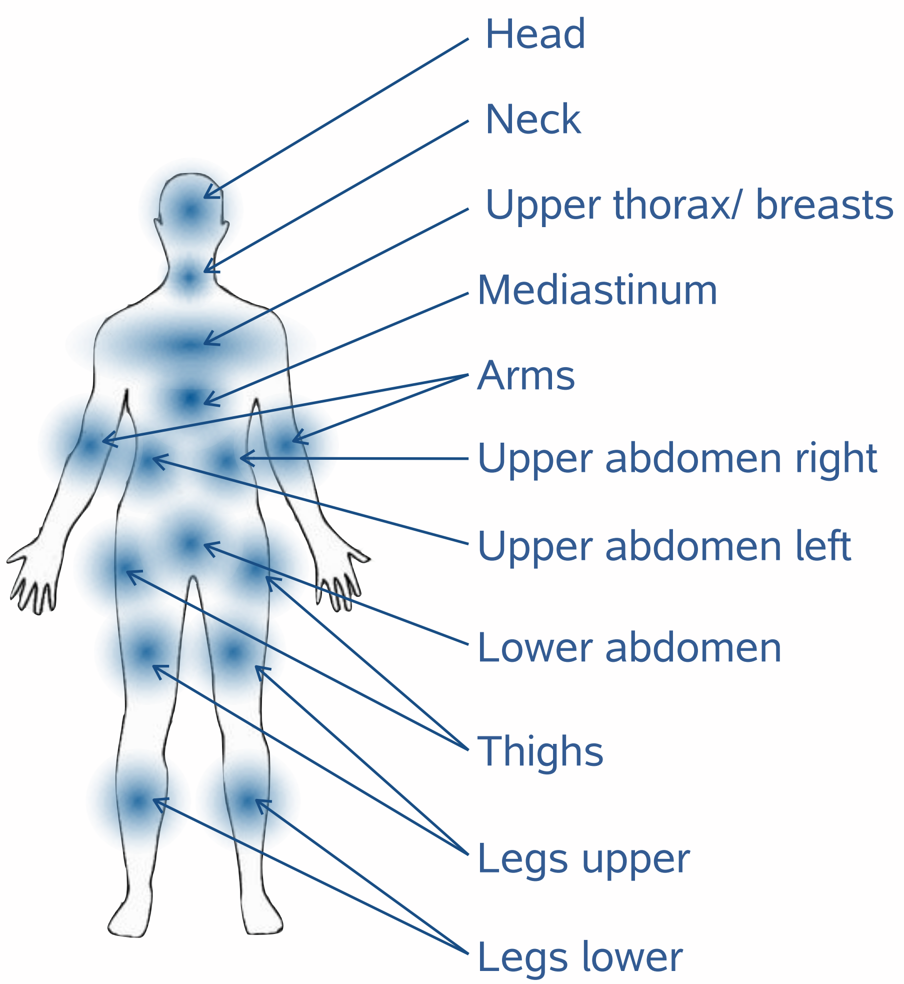
We find gender–specific clusters of varying diabetes prevalence

Gradient Maps

Gradient Heat Maps

Visualisation of network attention and respective analysis

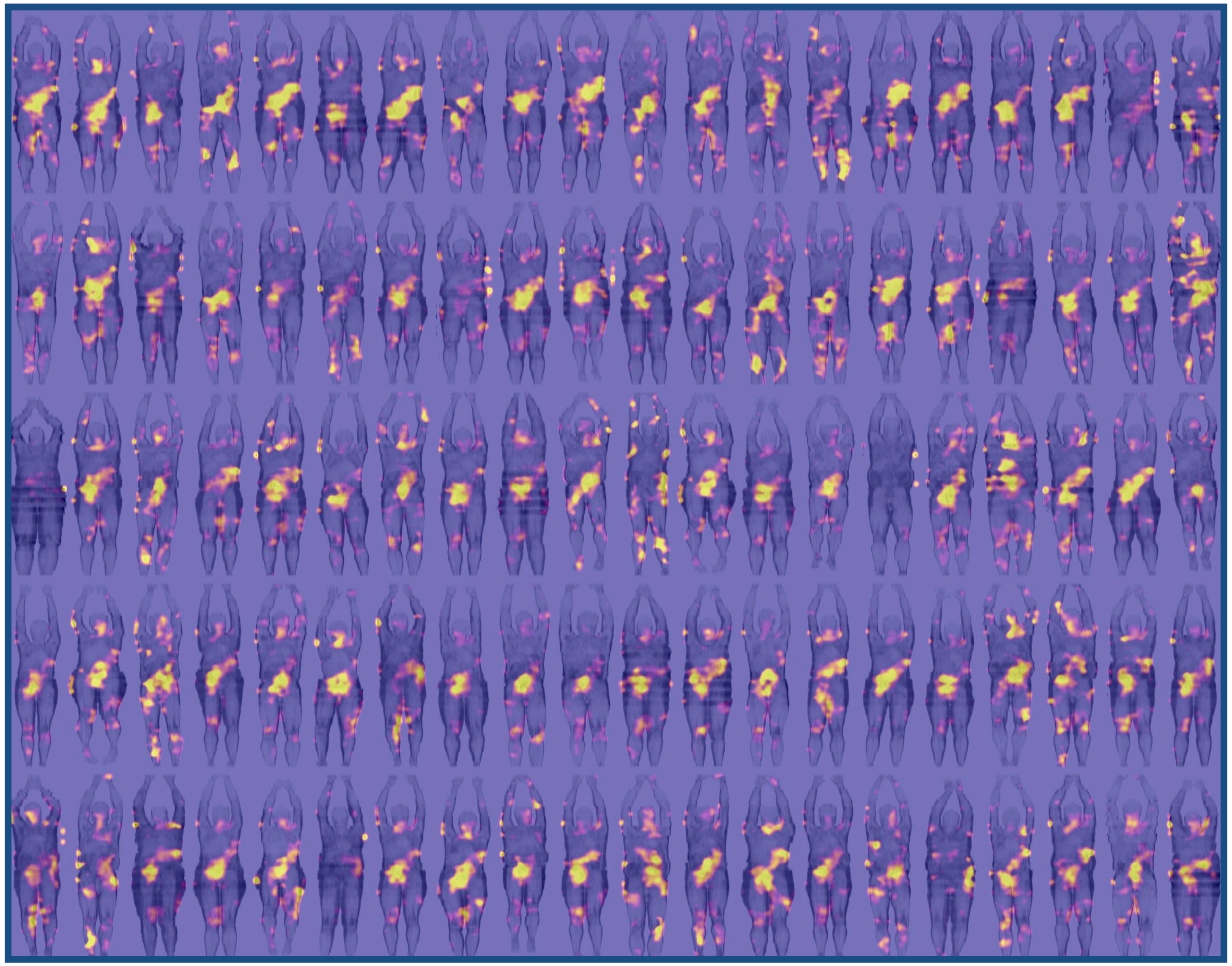


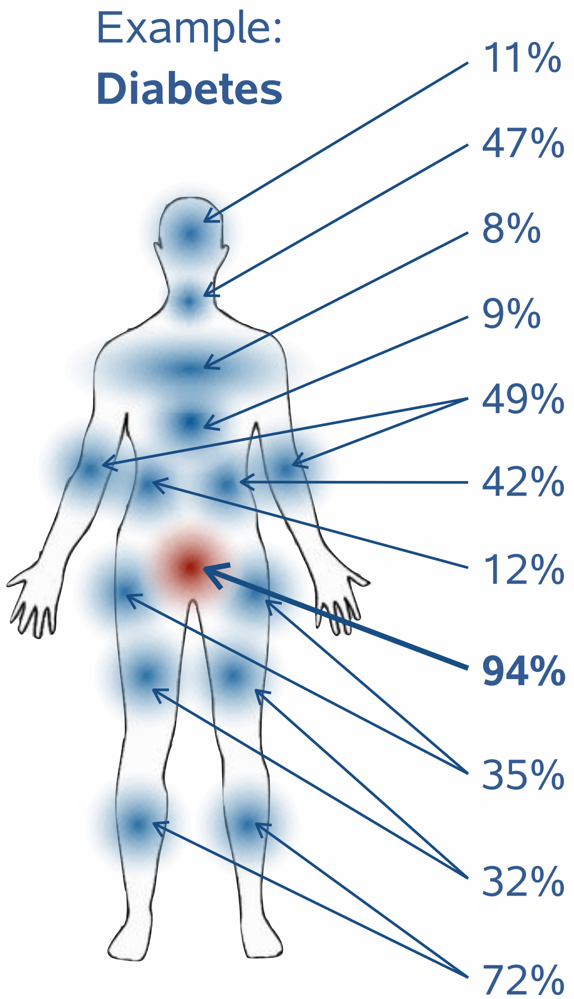


Medical experts classify the anatomical location of highlights to find trends

Gradient Heat Maps

Visualisation of network attention and respective analysis





Medical experts classify the anatomical location of highlights to find trends

Discussion

Discussion

Contribution

Deep learning analysis of full body MRT scans

Proof of concept for diabetes diagnosis from MRT scans

Respective benchmark scores (Diabetes AUC of ~0.7-0.8)

Analysis of representation space

Feature-specific heat maps

Limitations

Quantity of available data

Resolution of MRT scans (level of detail)

Non-representative sample distribution

Very few diabetes positives

Medical Application

Diabetes diagnosis solely based on HbA1c not very accurate

Oral Glucose testing requires prior fasting

Hence, interest for novel diabetes tests exists

MRT-based diagnosis could test for several diseases

Capacity and cost represent limiting factors

Future Work

Validate results on enhanced dataset

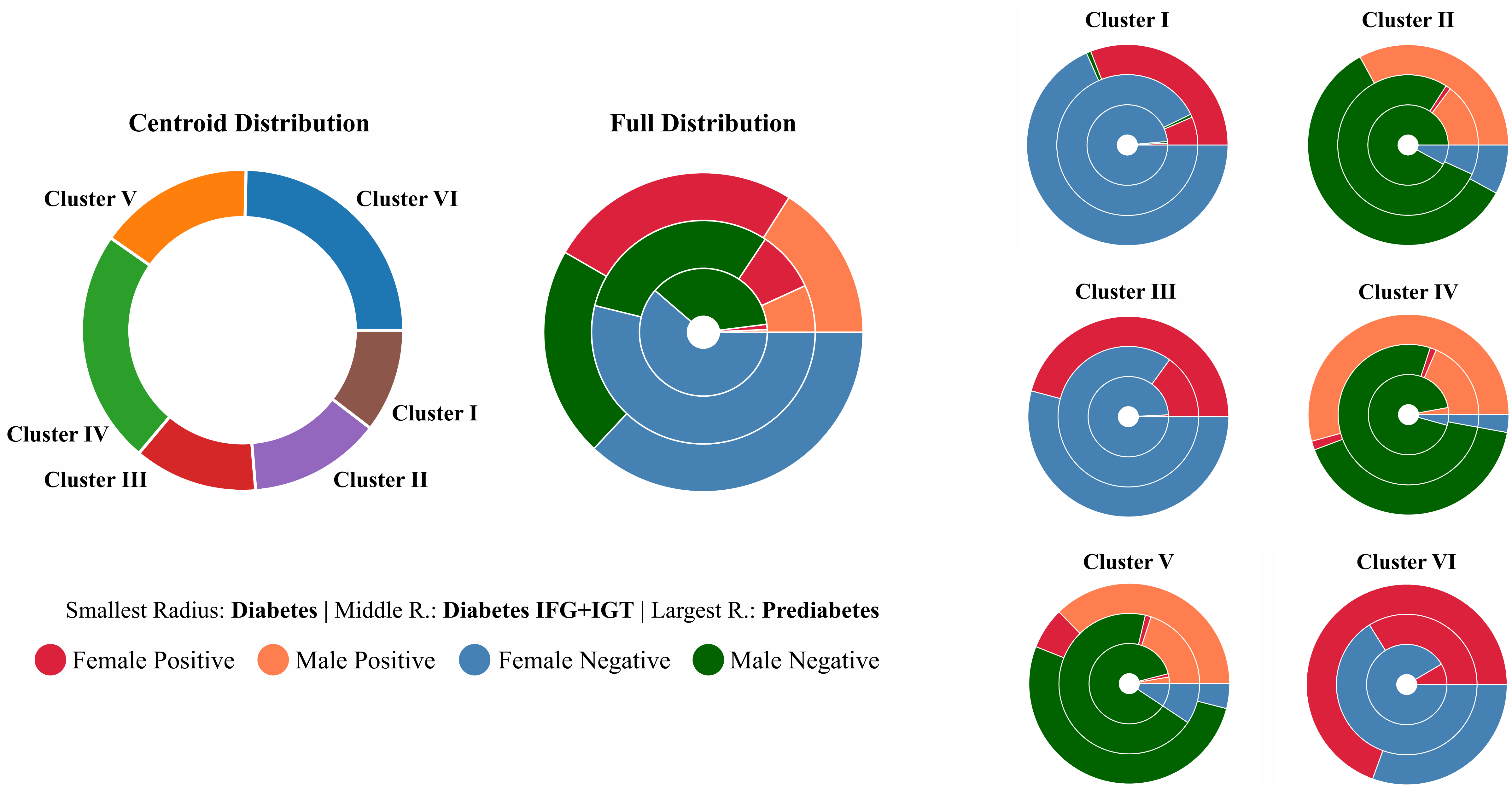
Take multiple scans per patient into consideration

Evaluate alternative approaches (segmentation, attention, ..)

Thank you!

Appendix

k–means Clustering



Performance Summary

Regression

- **Body Mass Index performs best**
- **Insulin Sensitivity and HbA1c are challenging**

	MAE
Age	2%
Body Mass Index	15%
HbA1c	45%
Insulin Sensitivity	

*

Performance Summary

Regression

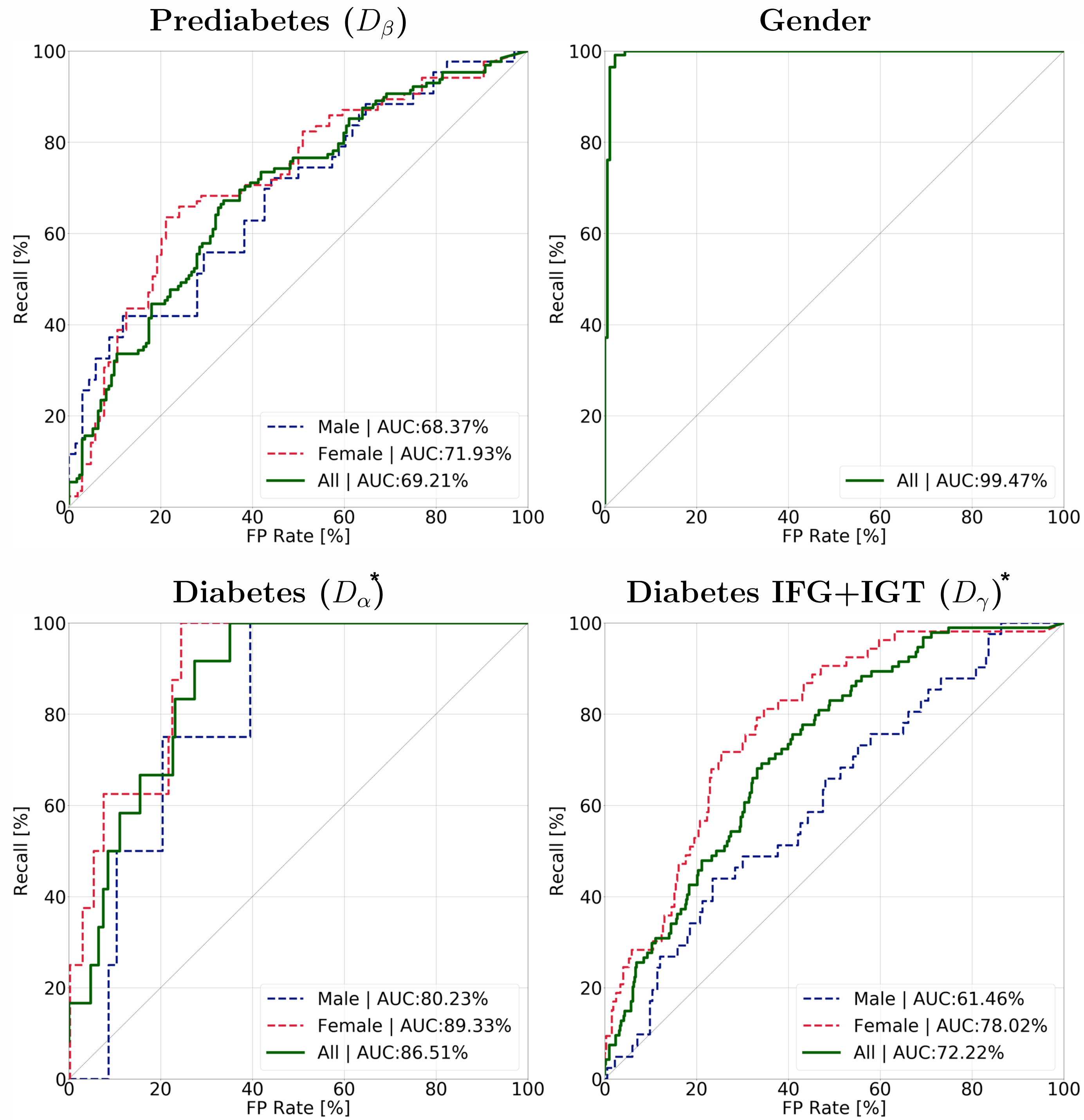
- Body Mass Index performs best
- Insulin Sensitivity and HbA1c are challenging

	MAE
Age	2%
Body Mass Index	15%
HbA1c	45%
Insulin Sensitivity	

Classification

- AUC ROC used as key metric
- Gender AUC converges to >99%
- Female AUC is higher for every diabetes label
- Diabetes AUC varies according to prevalence:

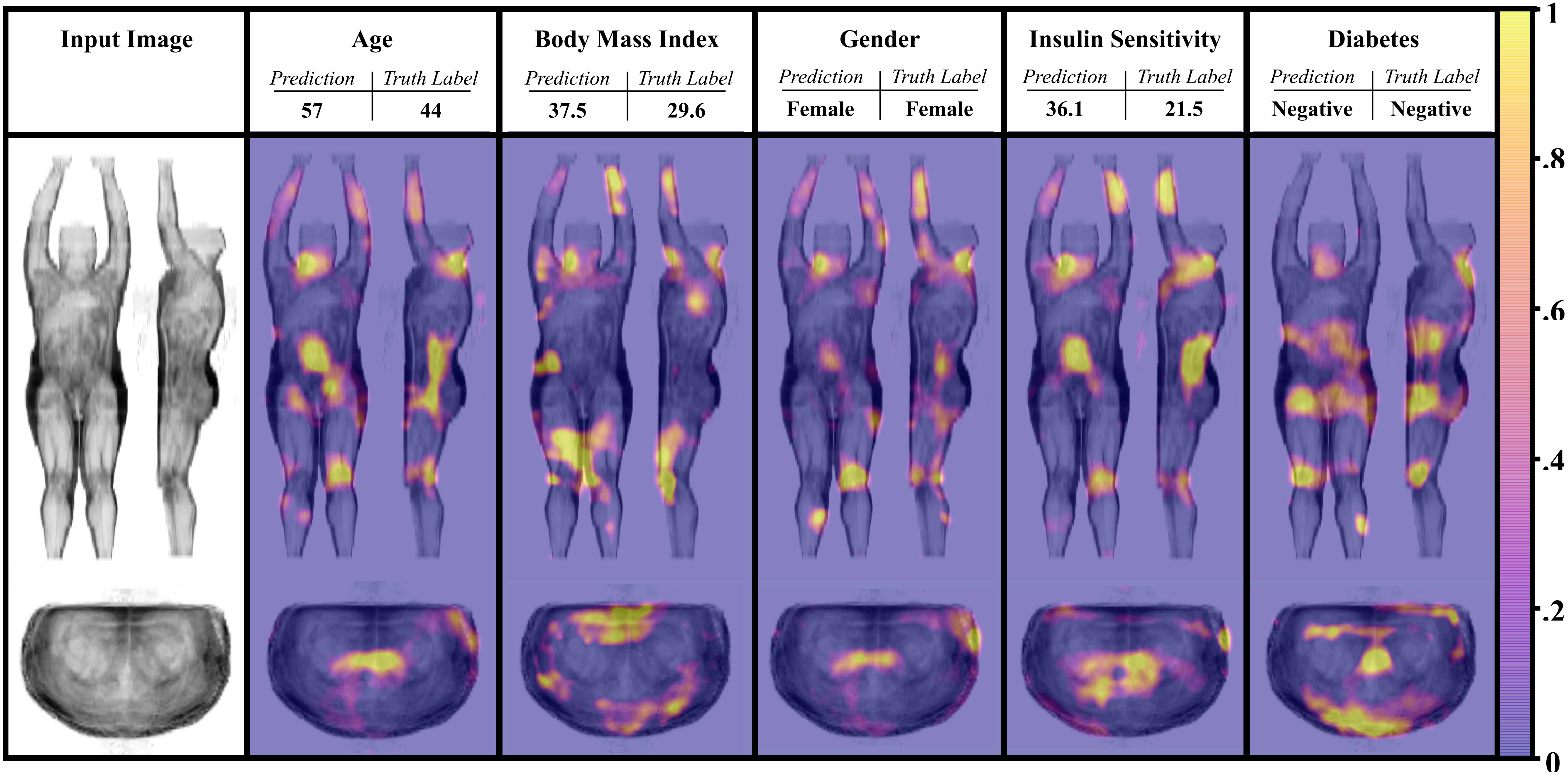
	Prevalence	AUC
Diabetes	2%	0.87
Diabetes IFG+IGT	15%	0.72
Prediabetes	45%	0.69



* Computed on test– and validation set due to critically low number of positives

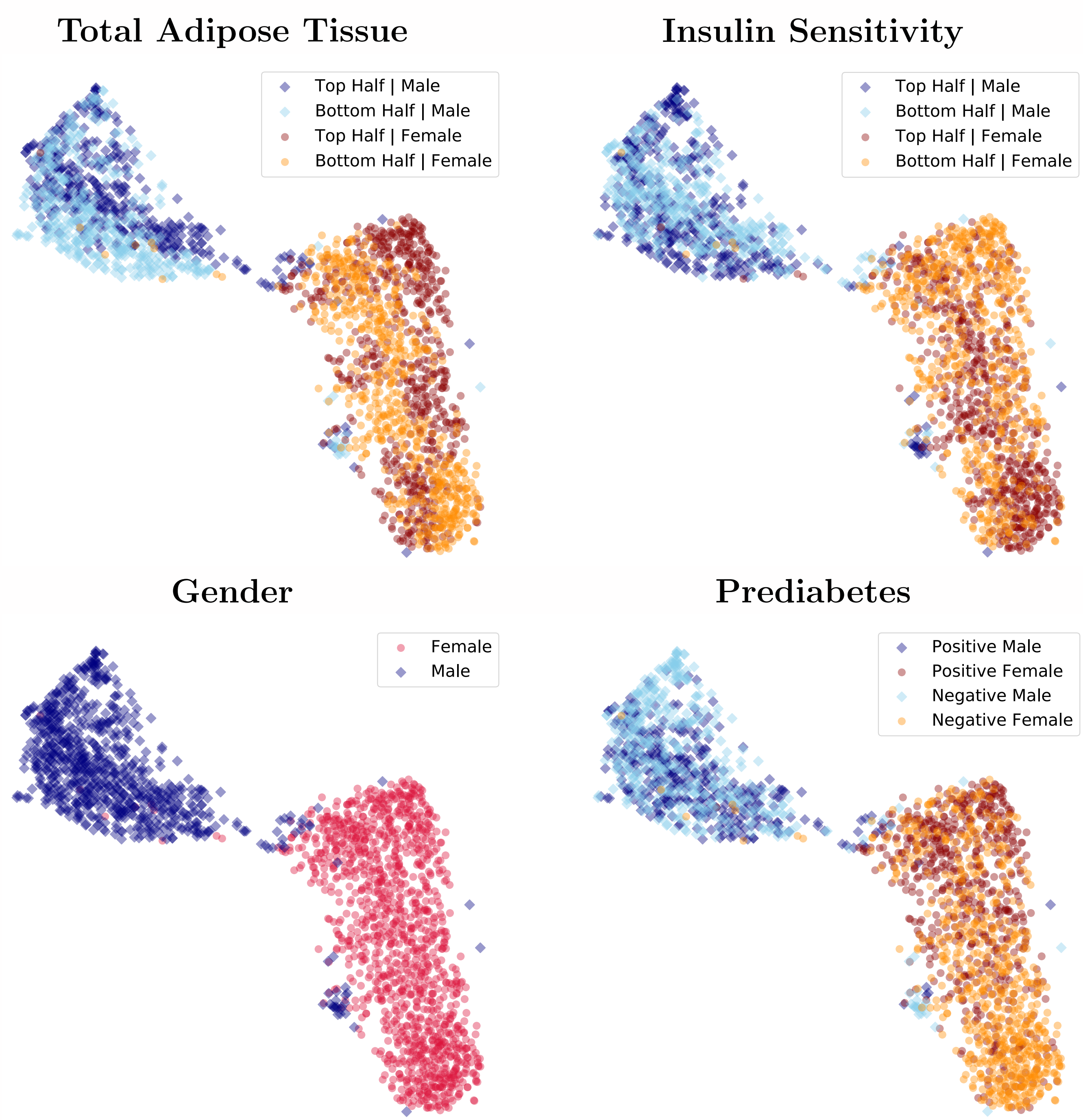
A Slide with Heat Map Examples

Dummy text



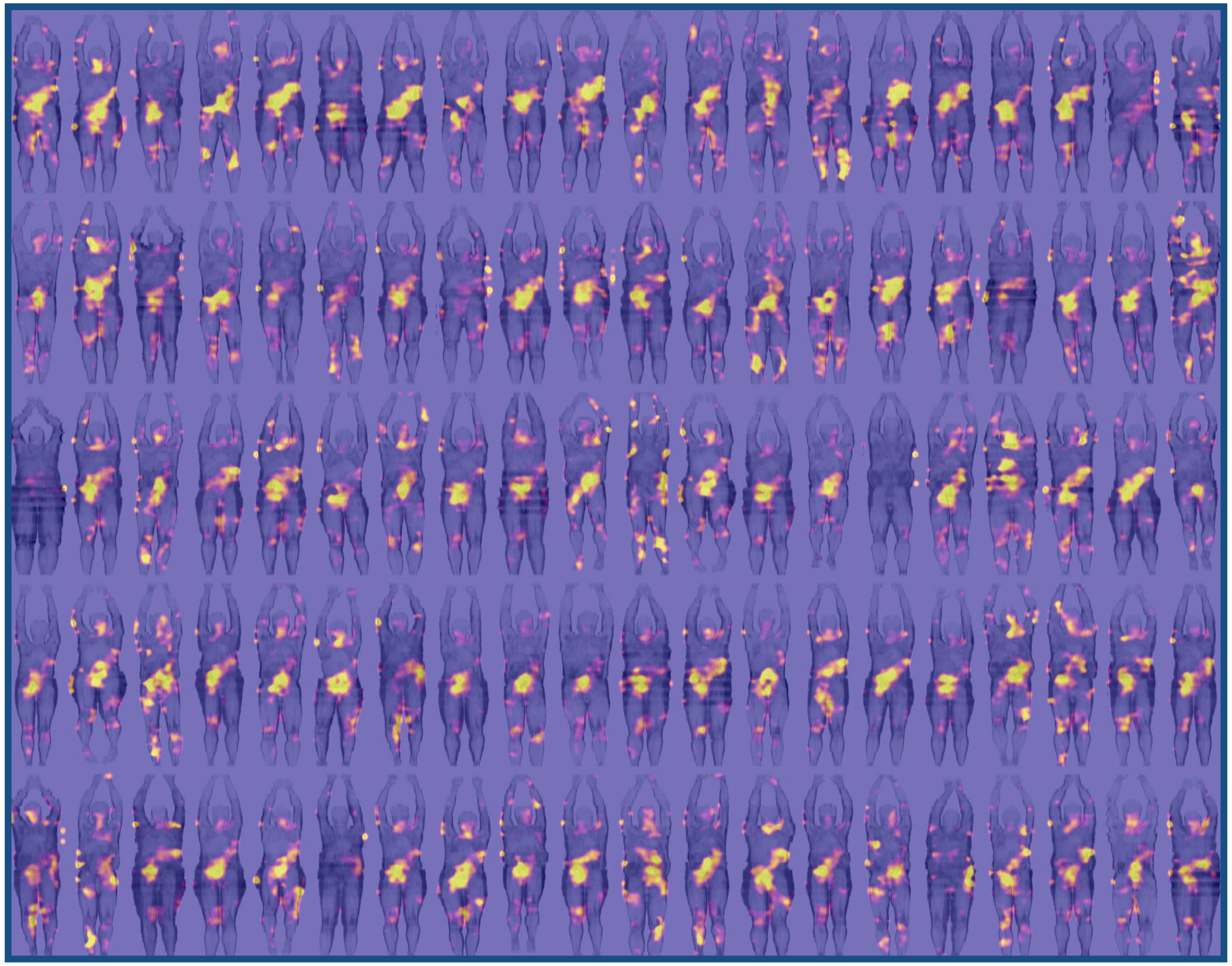
t–Distributed Stochastic Neighbour Embedding

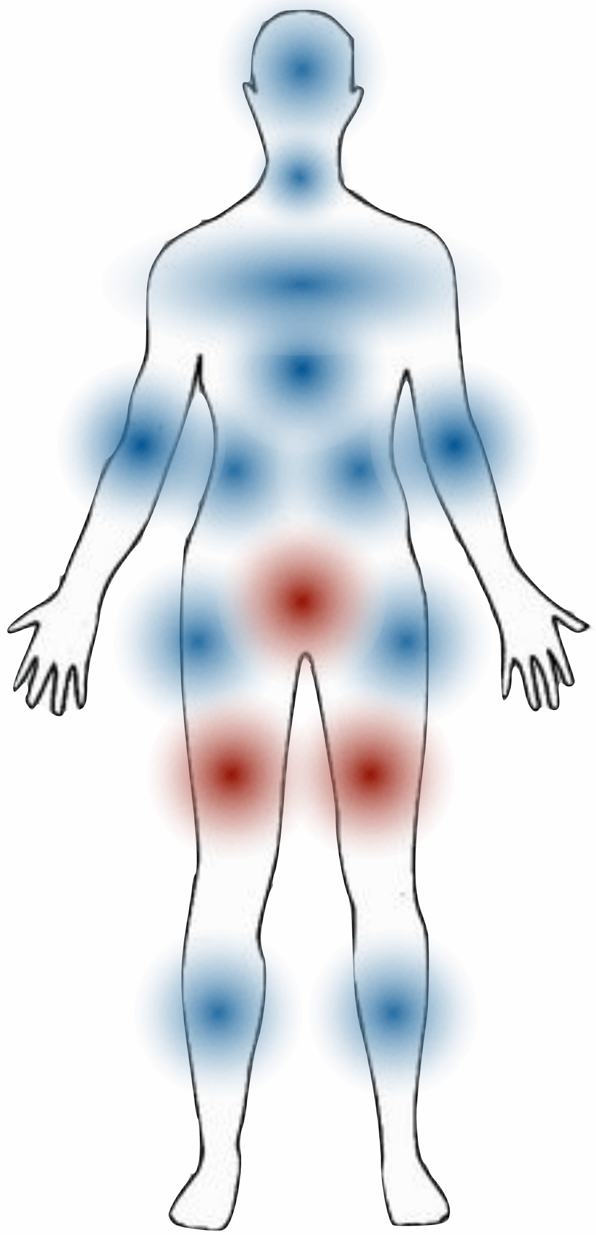
t–SNE maps the high–dimensional representation to 2D for visual analysis



A Slide with Heat Map Examples

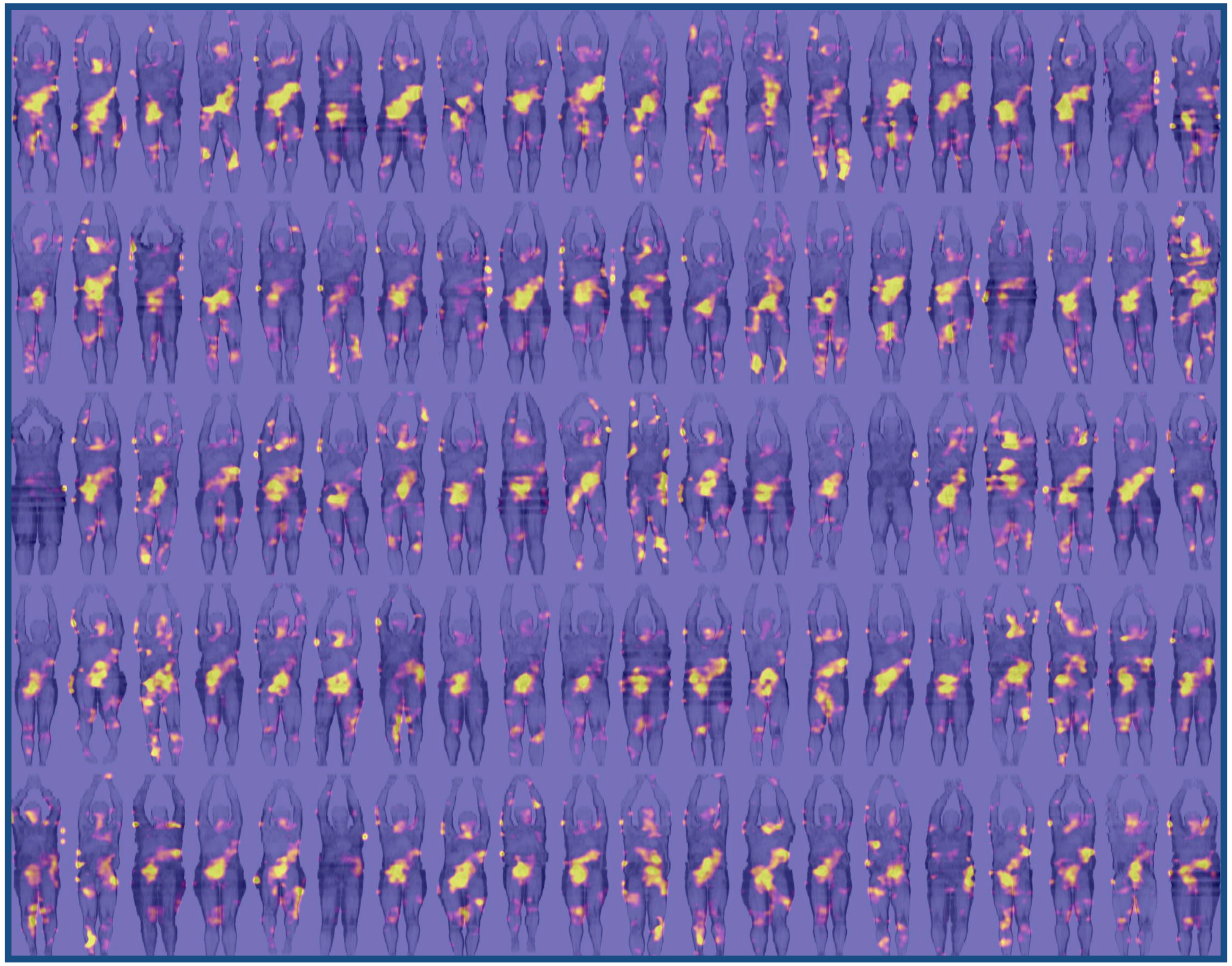
Diabetes

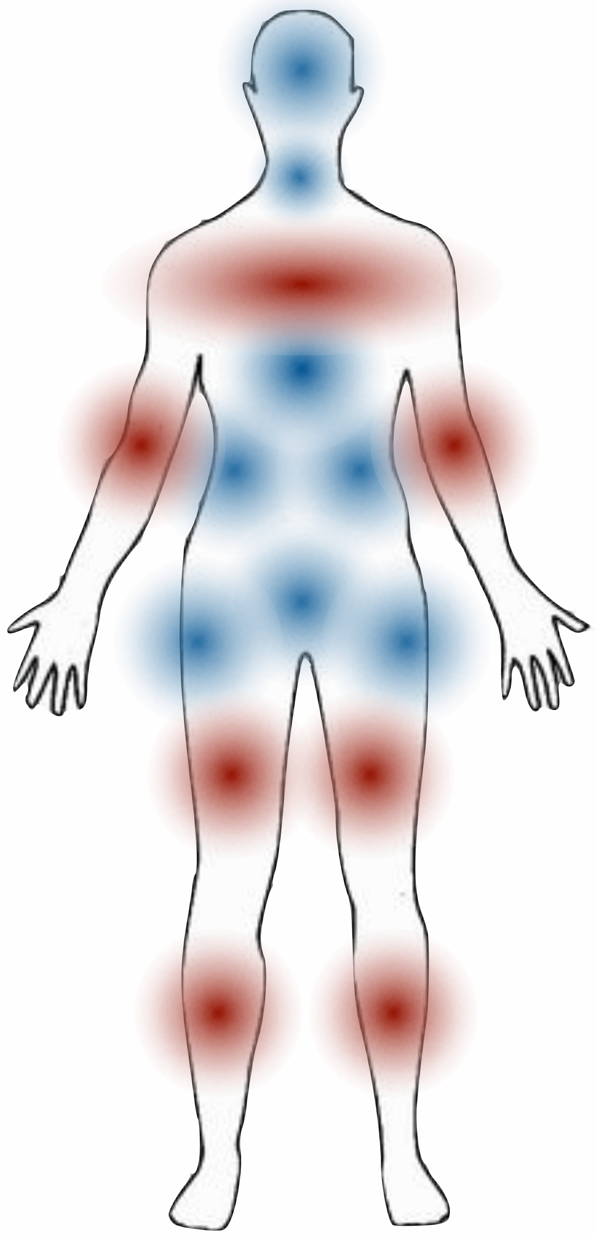




A Slide with Heat Map Examples

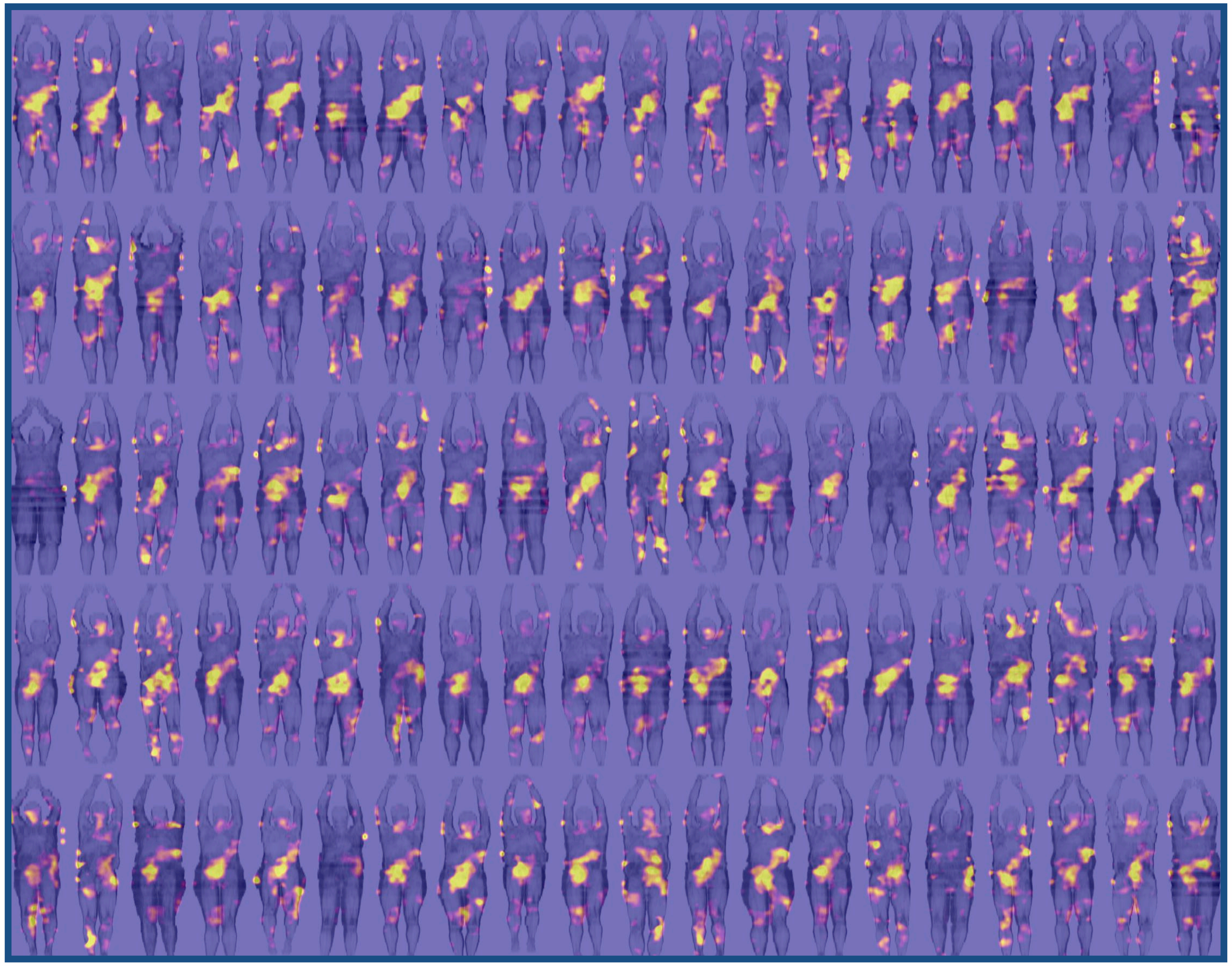
Gender

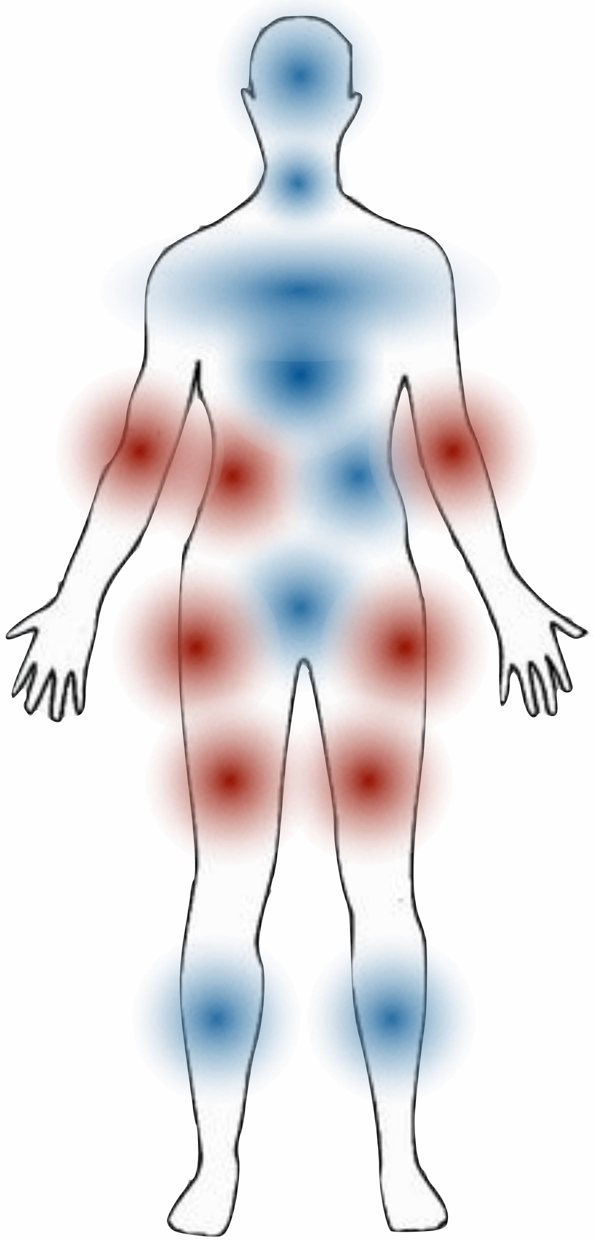




A Slide with Heat Map Examples

BMI





A Slide with Heat Map Examples

AGE