ETH zürich

Analysis of Electrocardiograms via Artificial Neural Networks For a Reliable Assessment of a Possible Myocardial Infarction

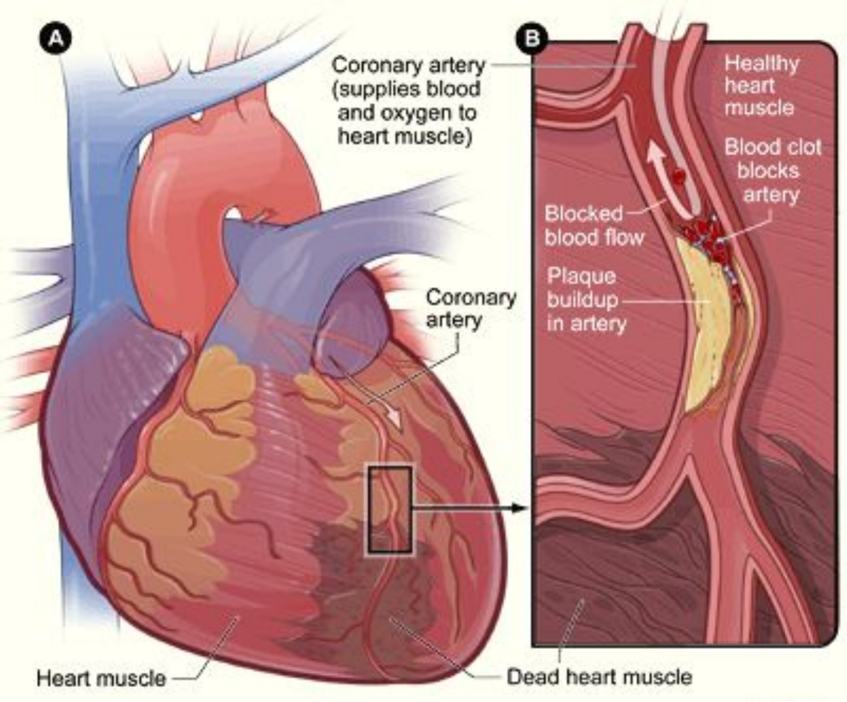
Alexey Gavryushin
Guglielmo Pacifico
Juan Fernandez de la Garza
Sunho Kim

Index

1.1 Problem Setting

What is Myocardial Infarction?

The presence of acute myocardial injury, detected by abnormal cardiac biomarkers in the setting of evidence of acute myocardial ischemia.



⇩

Happens when one or more areas of the heart muscle don’t get enough oxygen. This happens when blood flow to the heart muscle is blocked.

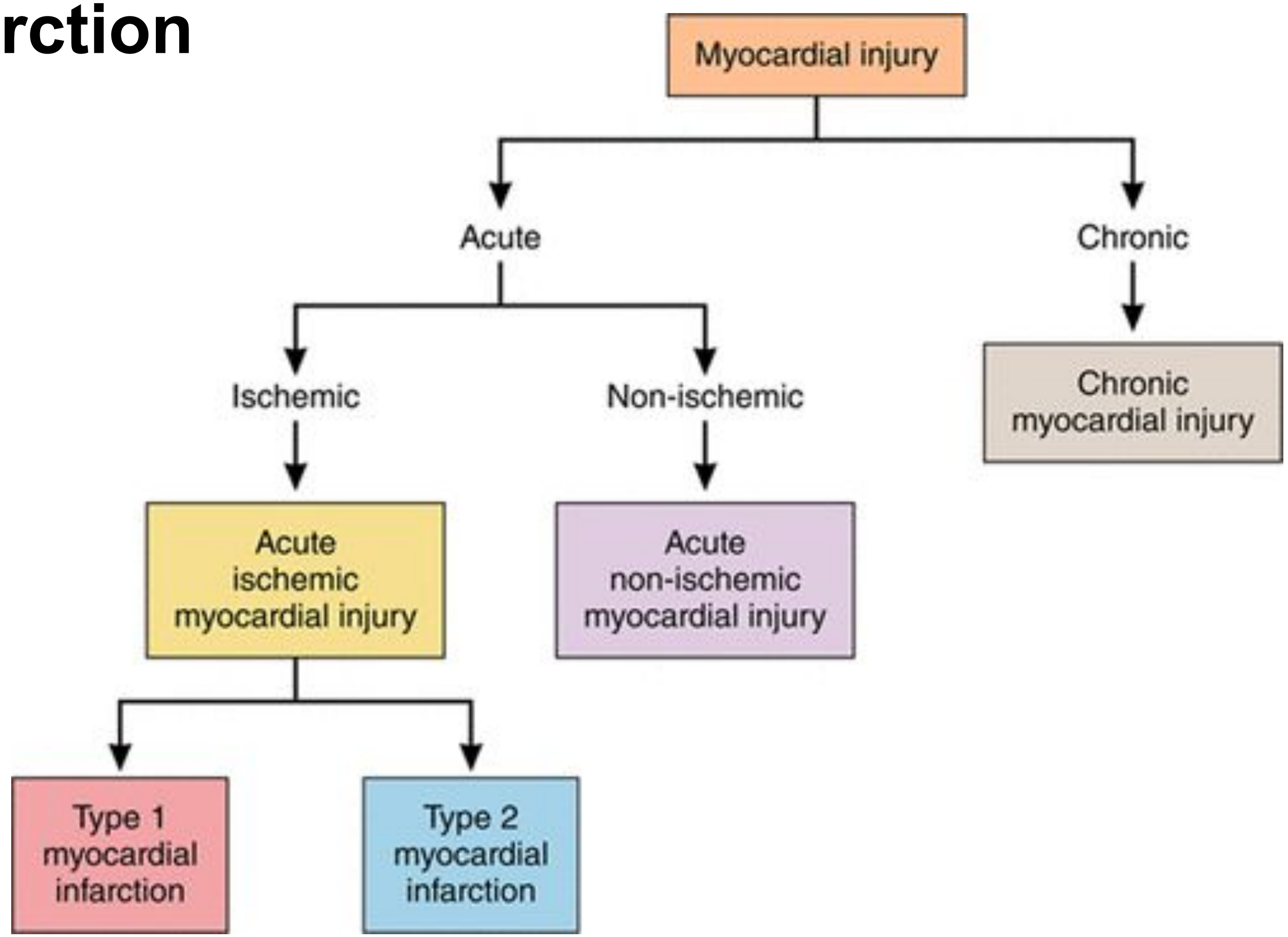
1) https://doi.org/10.1161/CIRCULATIONAHA.119.040631 Circulation. 2019;140:1661–1678
2) https://www.hopkinsmedicine.org/health/conditions-and-diseases/heart-attack

1.1 Problem Setting

Subtypes of Myocardial Infarction

Ischemic?

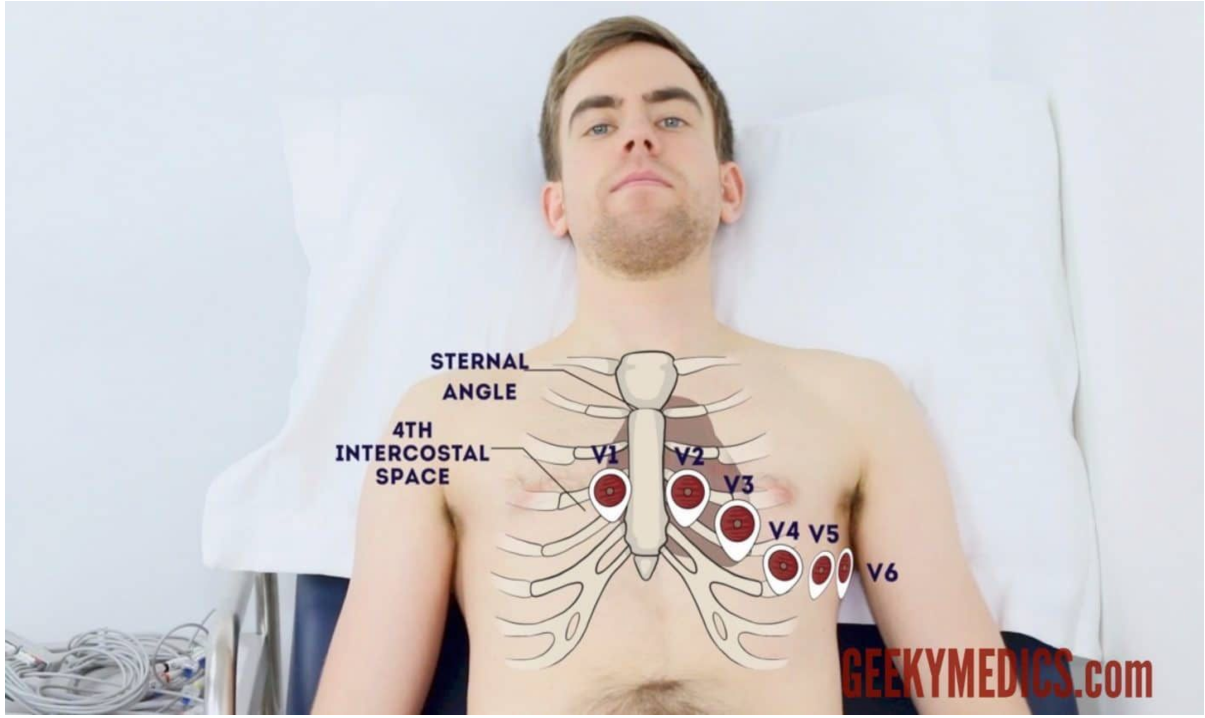
Deficient supply of blood to a body part that is due to obstruction of the inflow of arterial blood.

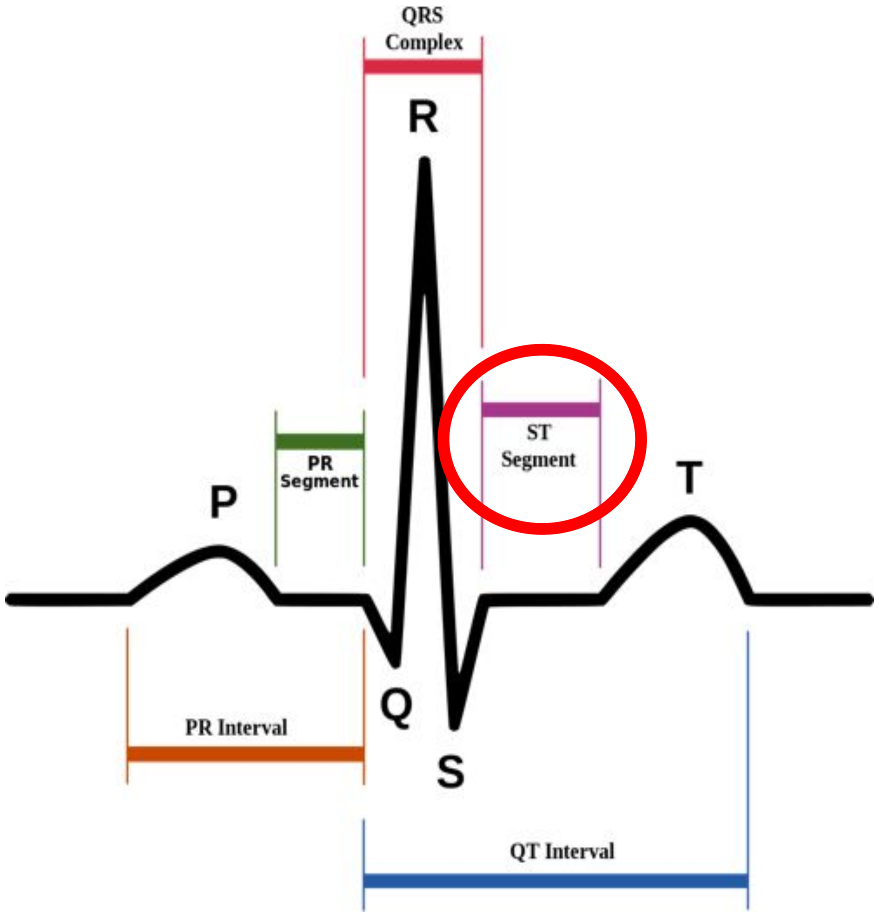


DeFilippis, A. P., Chapman, A. R., Mills, N. L., de Lemos, J. A., Arbab-Zadeh, A., Newby, L. K., & Morrow, D. A. (2019). Assessment and Treatment of Patients With Type 2 Myocardial Infarction and Acute Nonischemic Myocardial Injury. *Circulation*, *140*(20), 1661–1678. https://doi.org/10.1161/CIRCULATIONAHA.119.04063, Figure 1. Myocardial injury taxonomy.

1.1 Problem Setting

How can we detect MI? - ECG data





STEMI: ST-elevation myocardial infarction
NSTEMI: heart attack that usually happens when your heart needs oxygen can't be met

1.1 Problem Setting

How can we detect MI?

Figure 1) http://what-when-how.com/paramedic-care/the-monitoring-ecg-clinical-essentials-paramedic-care-part-1
Figure 2) https://www.samsung.com/global/galaxy/galaxy-watch4/
Figure 3) https://www.apple.com/uk/newsroom/2019/03/ecg-app-and-irregular-rhythm-notification-on-apple-watch-available-today-across-europe-and-hong-kong/

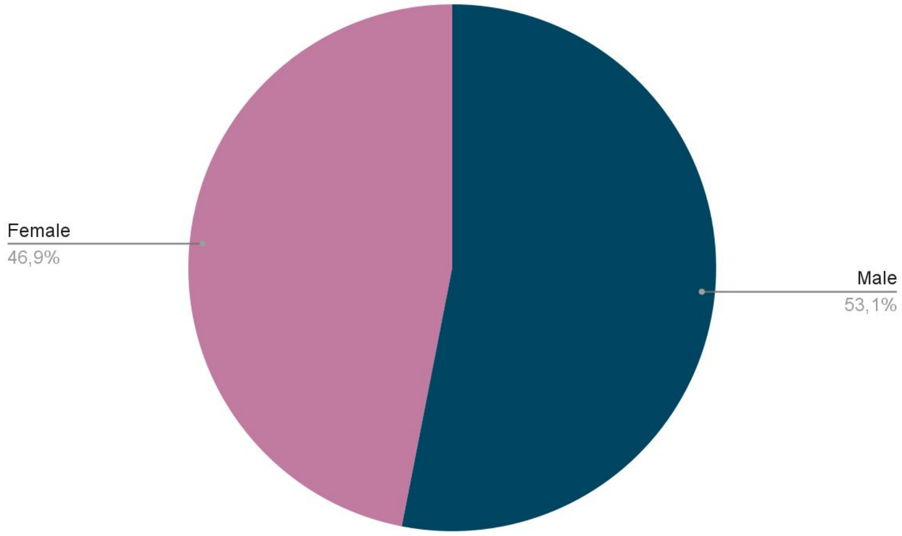
1.2 Project Idea

Why do we need a reliable ECG assessment?

"In Tanzania, I was involved in [training and equipping taxi drivers](https://www.hmpgloballearningnetwork.com/site/emsworld/article/12101337/the-struggles-of-ems-in-other-countries) in the place of ambulances and using motorcycle taxi riders as first responders."

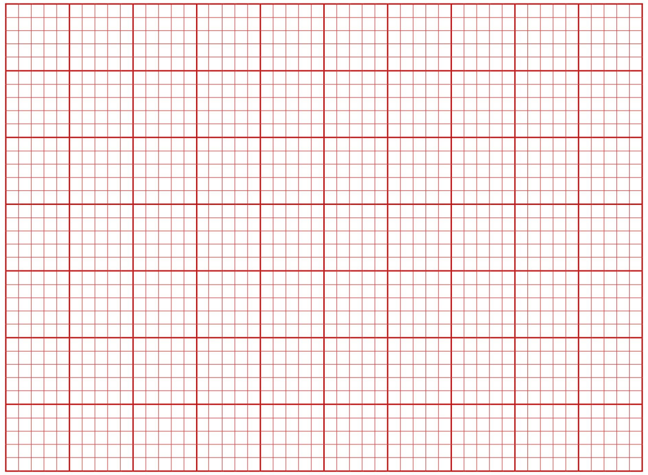
Figure 1) https://www.aljazeera.com/gallery/2019/4/1/ugandas-bicycle-ambulances-help-the-pregnant-sick-and-injured
https://www.hmpgloballearningnetwork.com/site/emsworld/article/12101337/the-struggles-of-ems-in-other-countries

2.1 Datasets

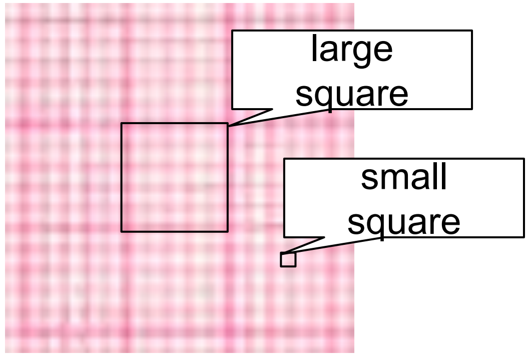


- PTB-XL ECG dataset
  - 21837 clinical 12-lead ECGs
  - 18885 patients, between 0 and 95 years old
  - 10 second length
  - 71 total different ECG statements conform to the SCP-ECG s

- Generate new “**PTB-V**” (“PTB-Visualized”) dataset
  - From time series to images
  - Python library “PIL” (Pillow) and the classes `Image` and `ImageDraw`



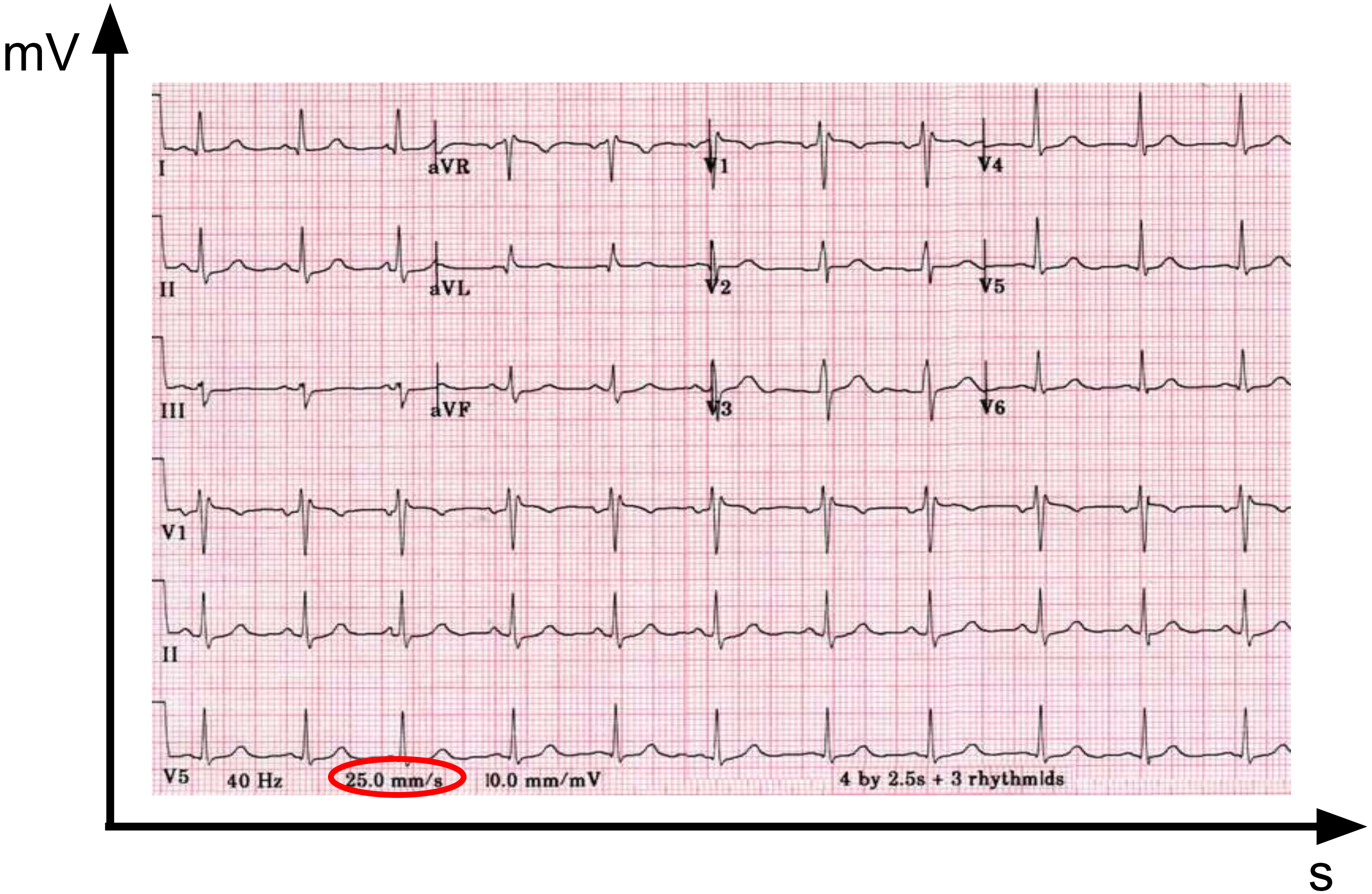
2.2 How to Read an ECG

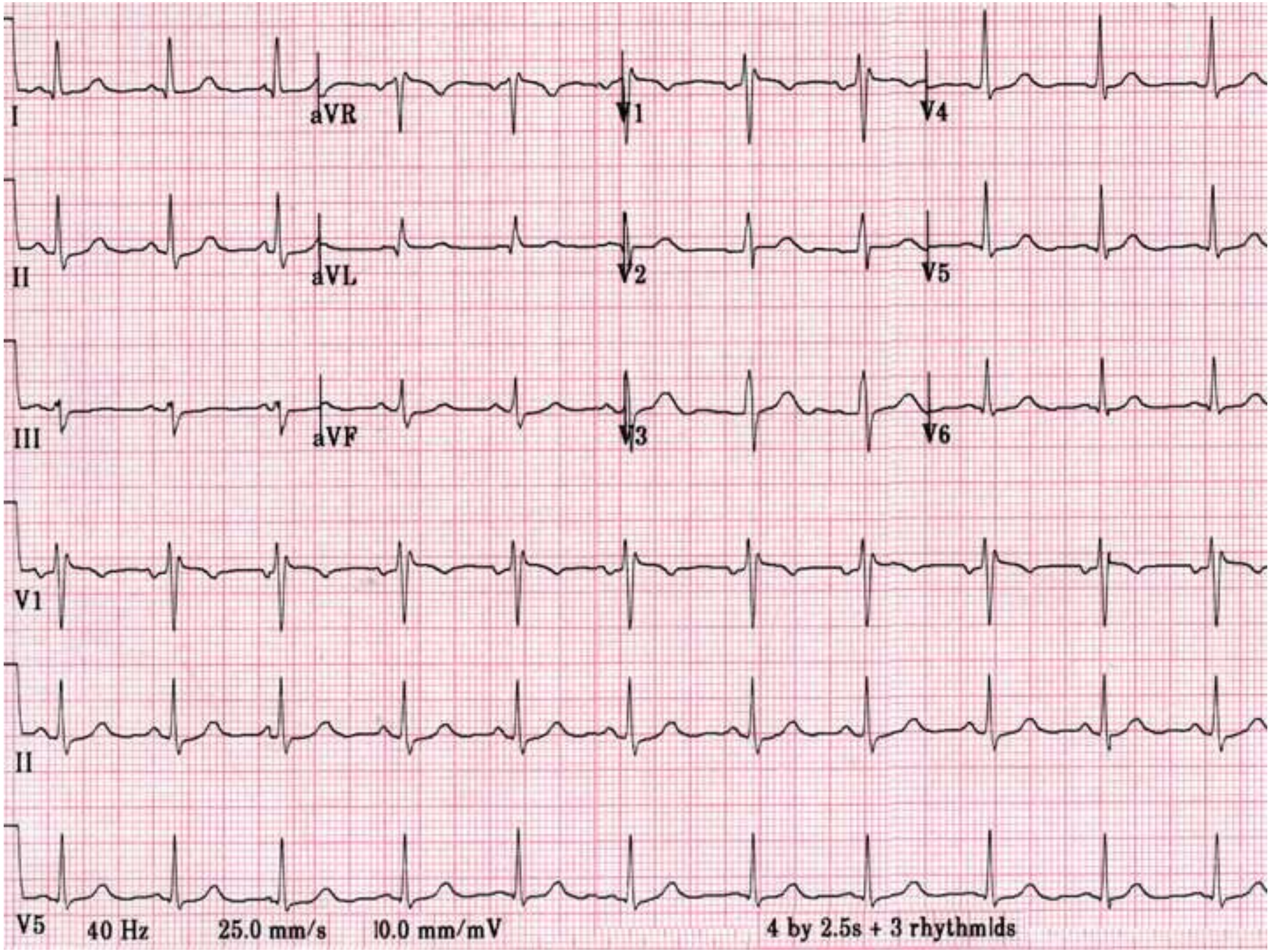


Default paper speed of 25mm/s:

- Each small square is 1 mm^2
- Each small square represents 0.04 s
- Each large square represents 0.2 s
- 5 large squares = 1 s
- 10 squares high = 1 mV

- **Chest leads**
  - V1: septal view of the heart
  - V2: septal view of the heart
  - V3: anterior view of the heart
  - V4: anterior view of the heart
  - V5: lateral view of the heart
  - V6: lateral view of the heart
- **Other leads**
  - Lead I: lateral view
  - Lead II: inferior view
  - Lead III: inferior view
  - aVR: lateral view
  - aVL: lateral view
  - aVF: inferior view

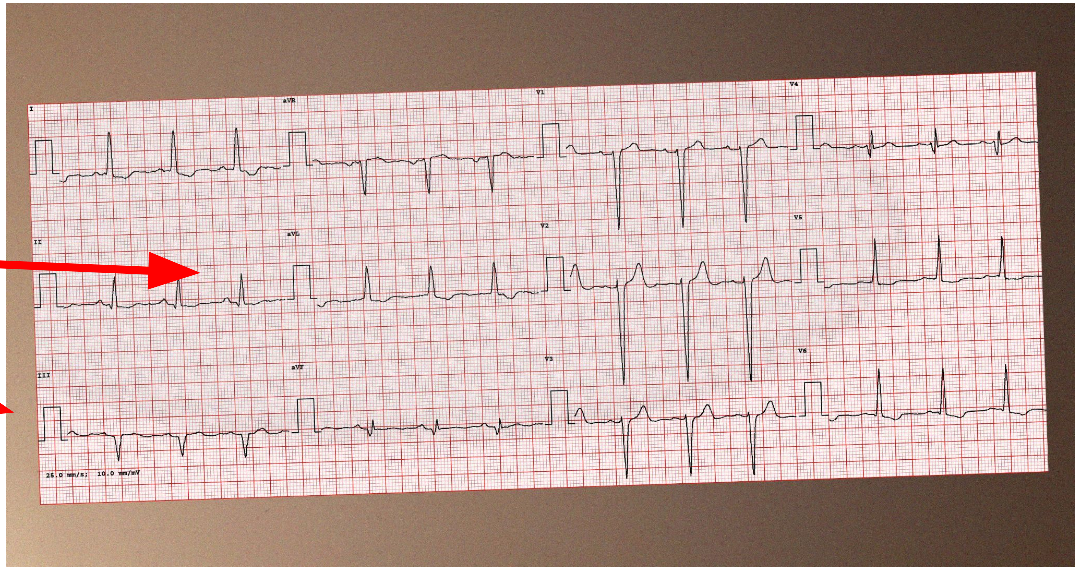


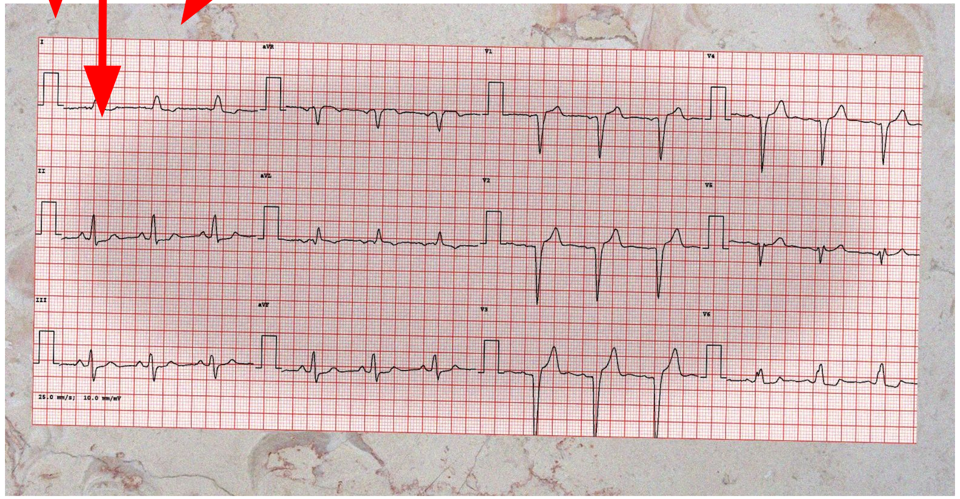

I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
V1
II
V5
40 Hz
25.0 mm/s
10.0 mm/mV
4 by 2.5s + 3 rhythmlds

2.3 Data augmentation

Goal : making the segmentation network more robust to different kinds of inputs

- shearing
- rotation
- blurring
- noise
- shadows and flashes
- different backgrounds

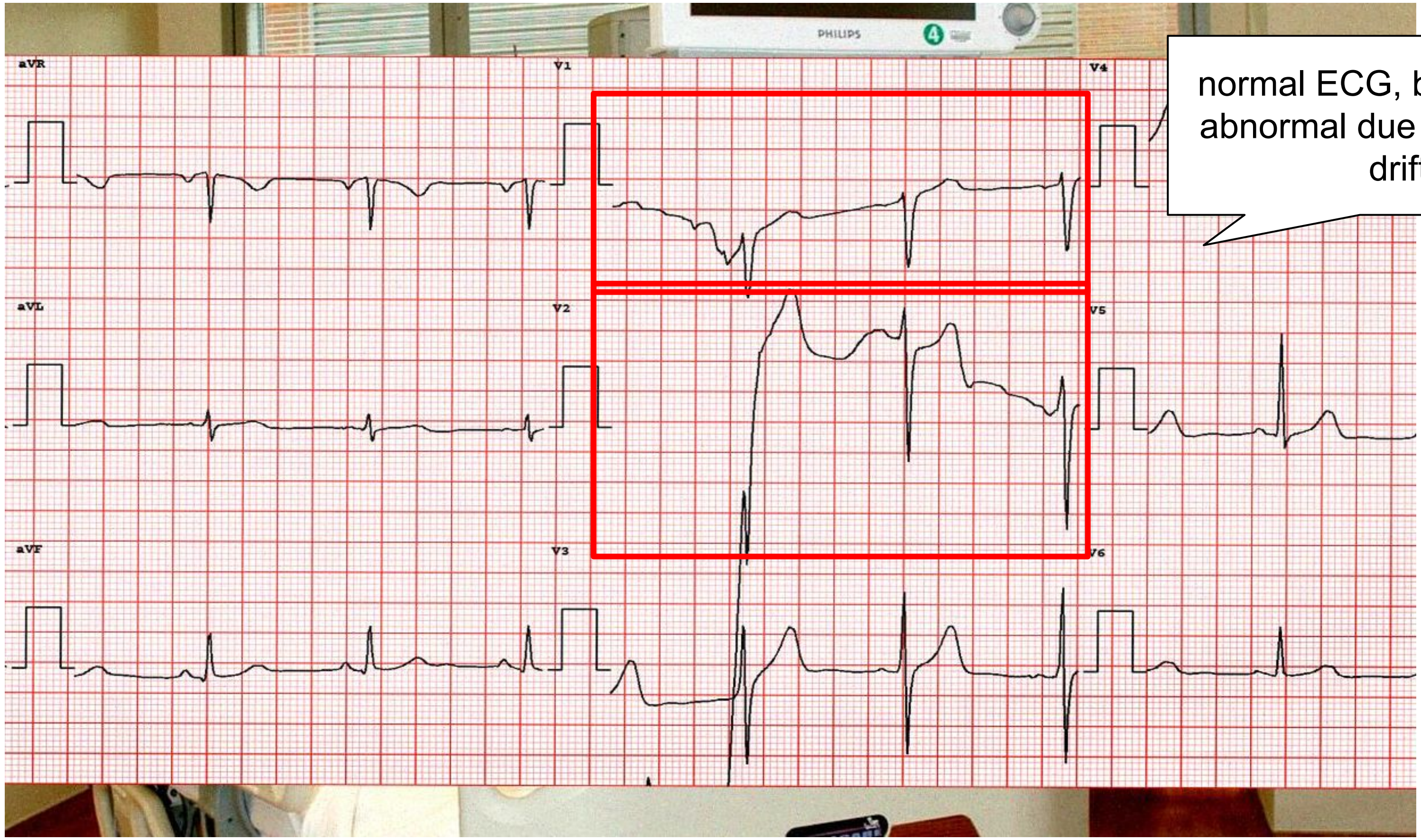




→ as a result we also generate more pictures

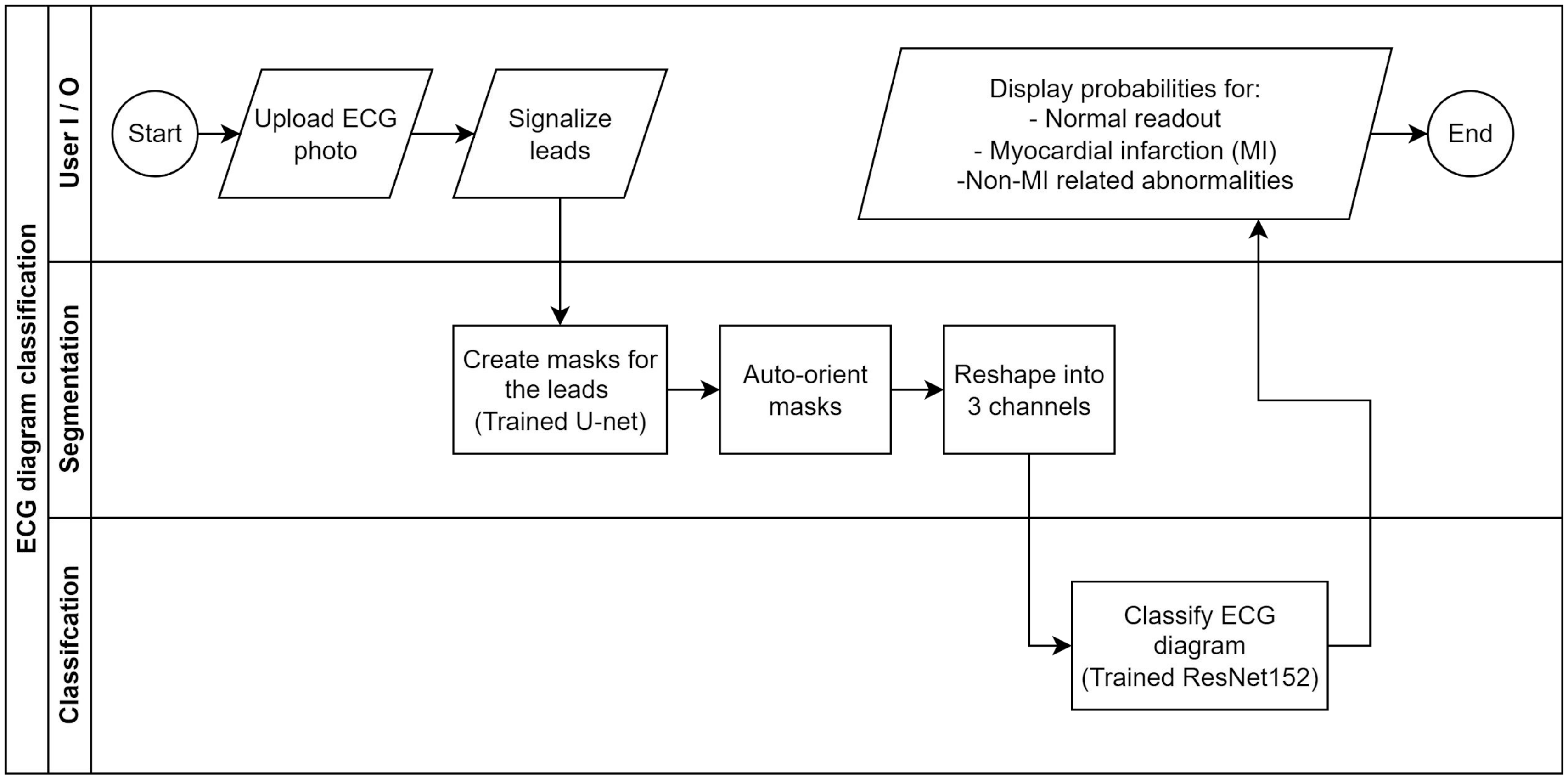
2.4 Problems with the data

- **Baseline drift**: Patient moving during the measurement

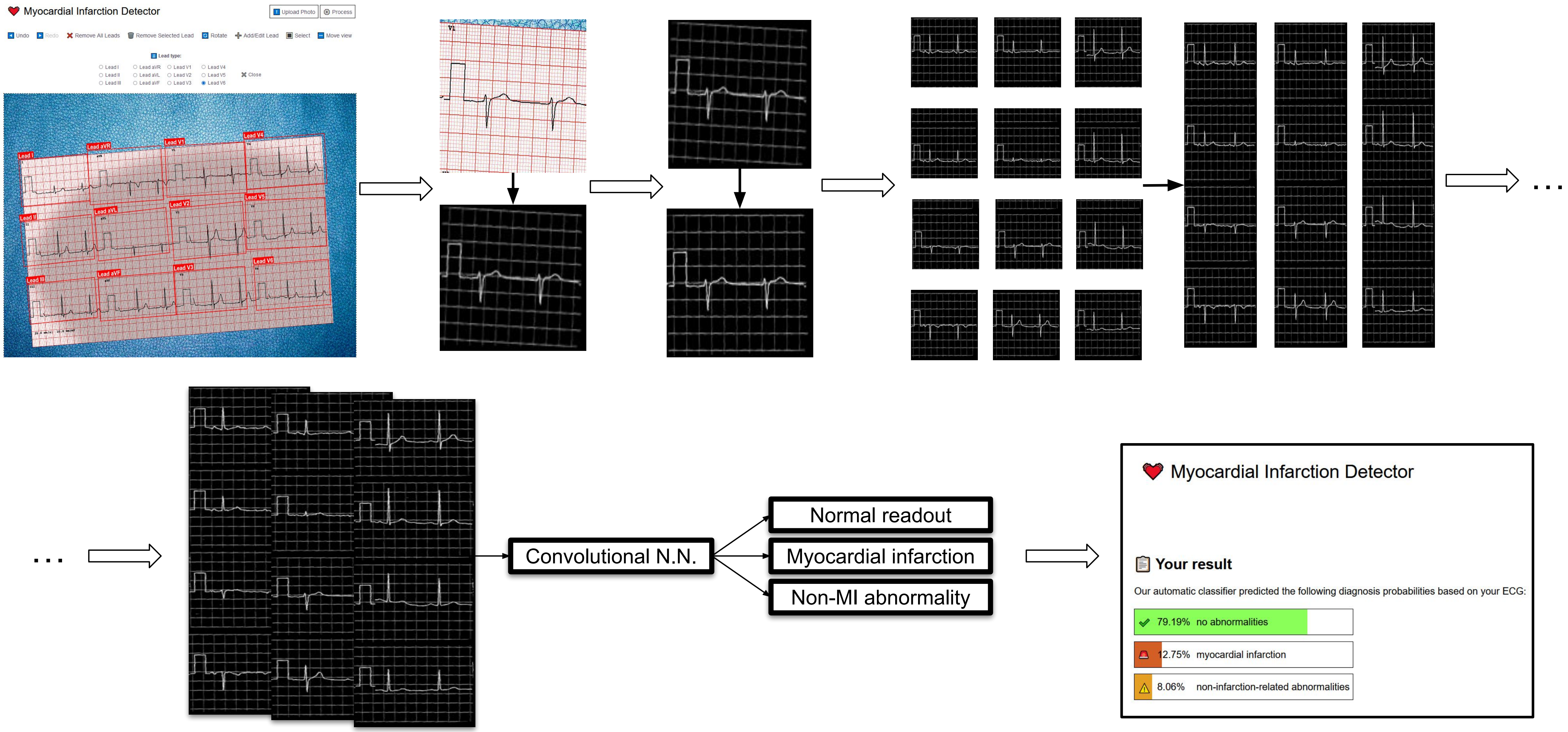


normal ECG, but appears abnormal due to baseline drift

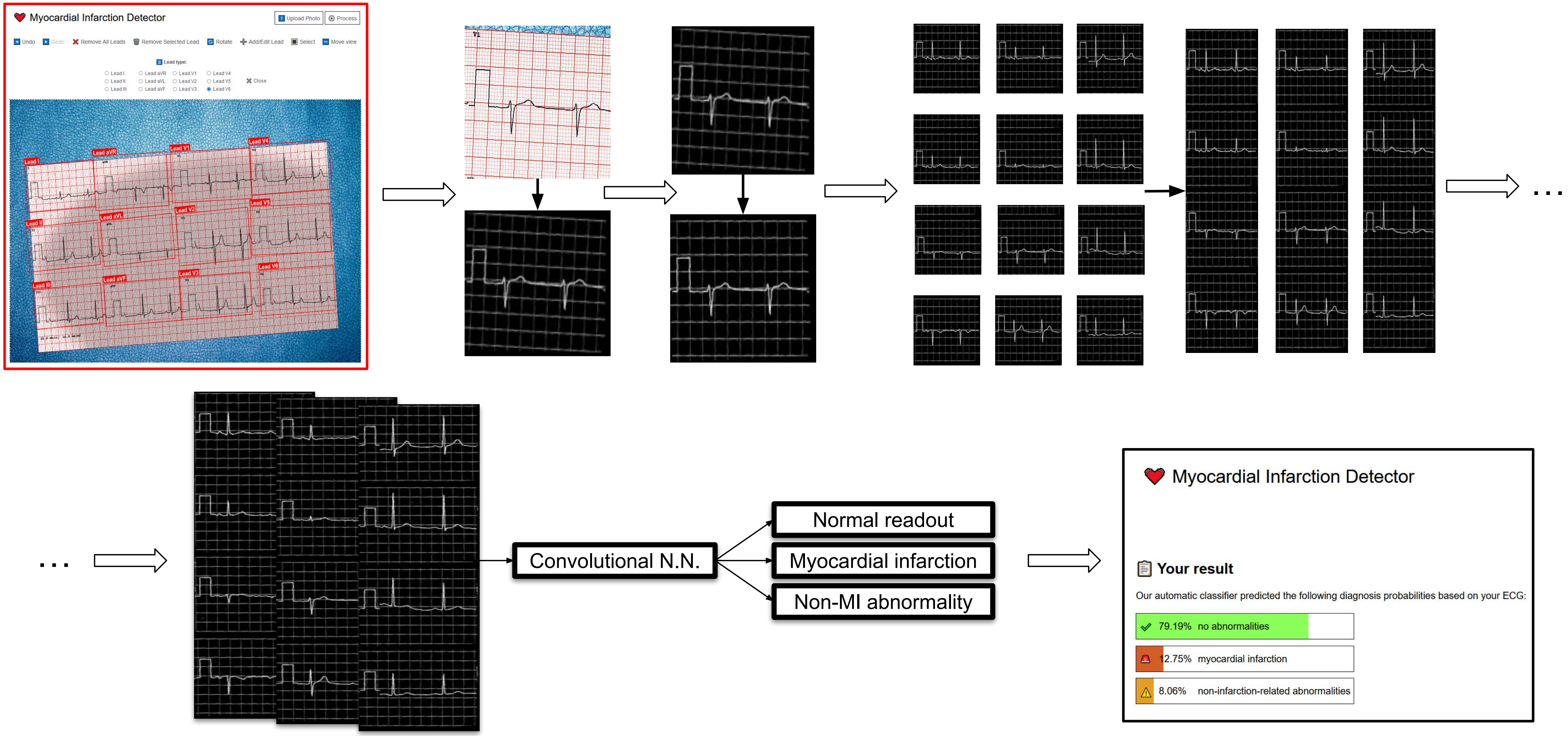
3 ECG Processing Overview



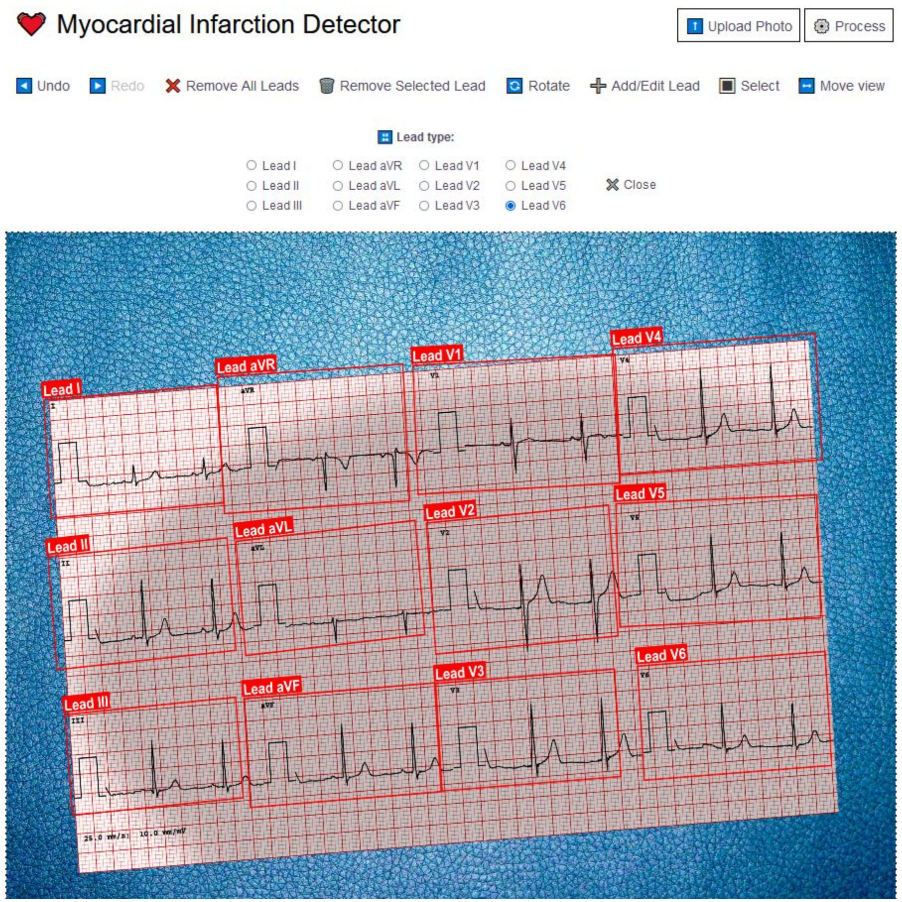
3 ECG Processing Overview

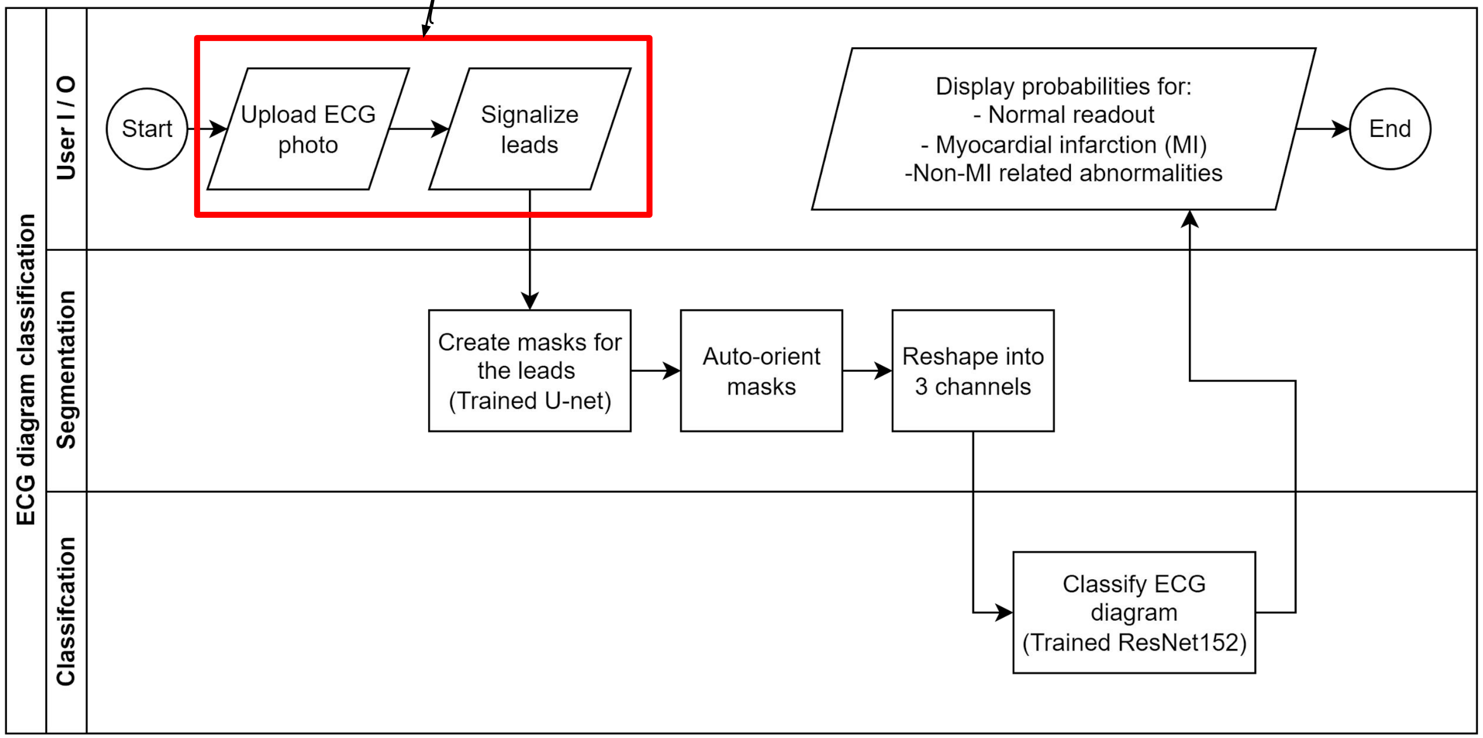


3 ECG Processing Overview

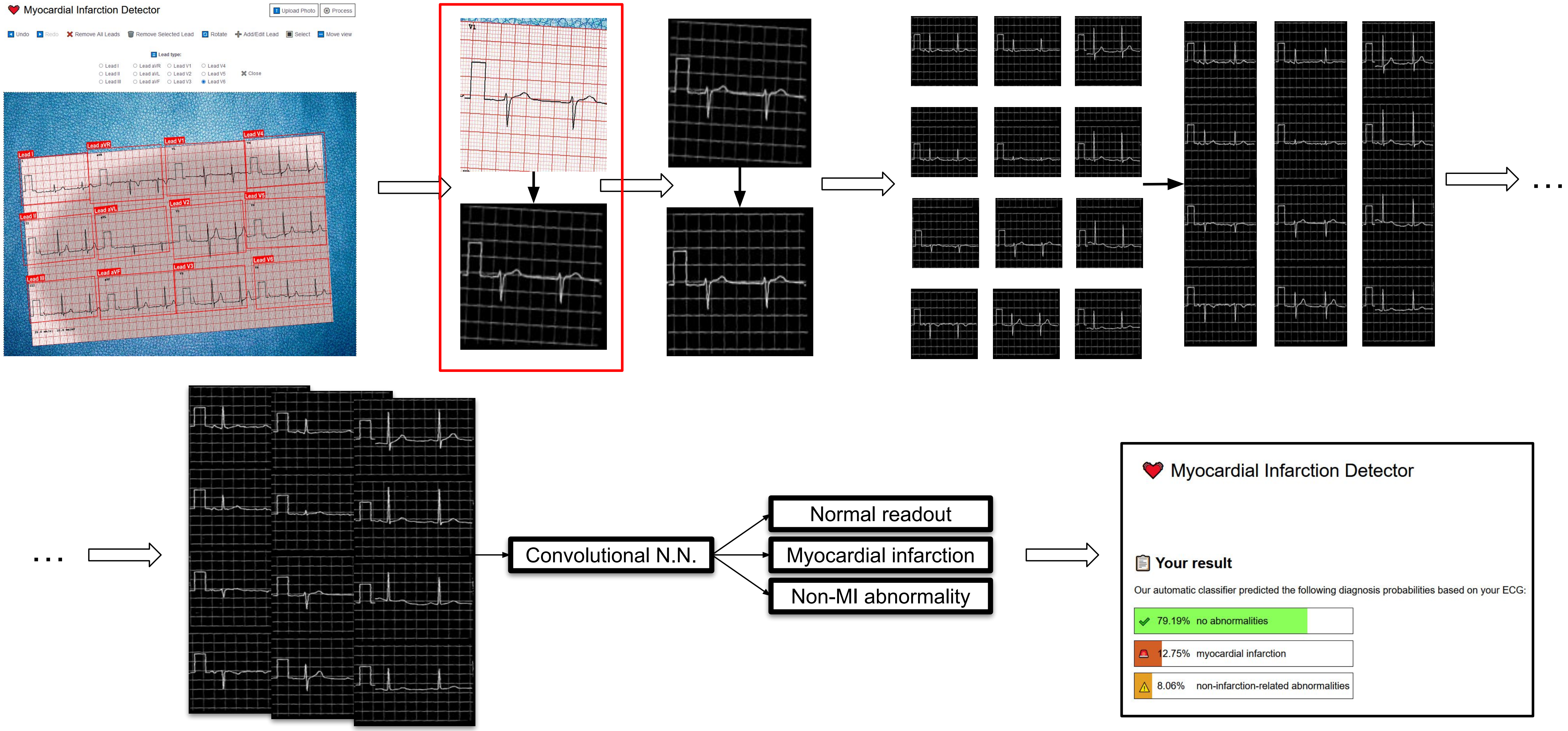


3 ECG Processing Overview

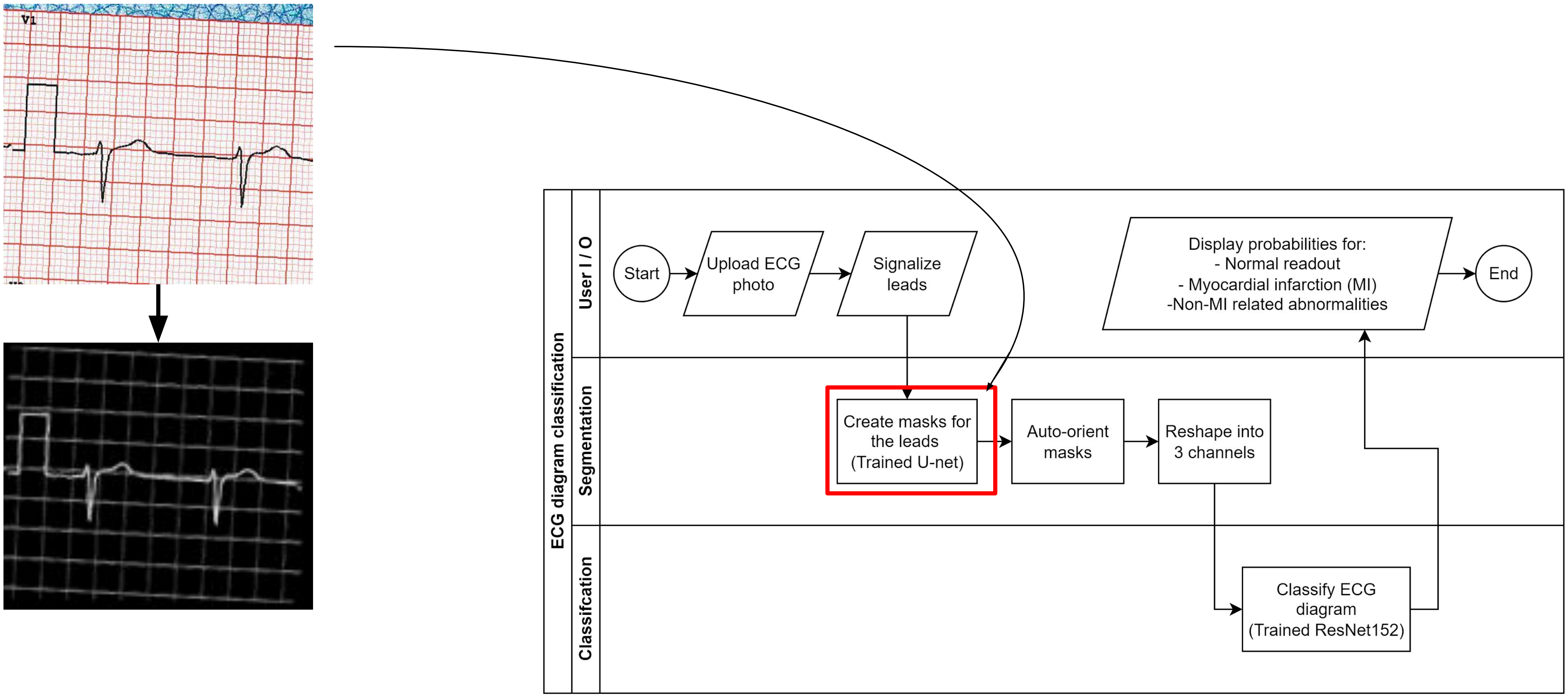




3 ECG Processing Overview



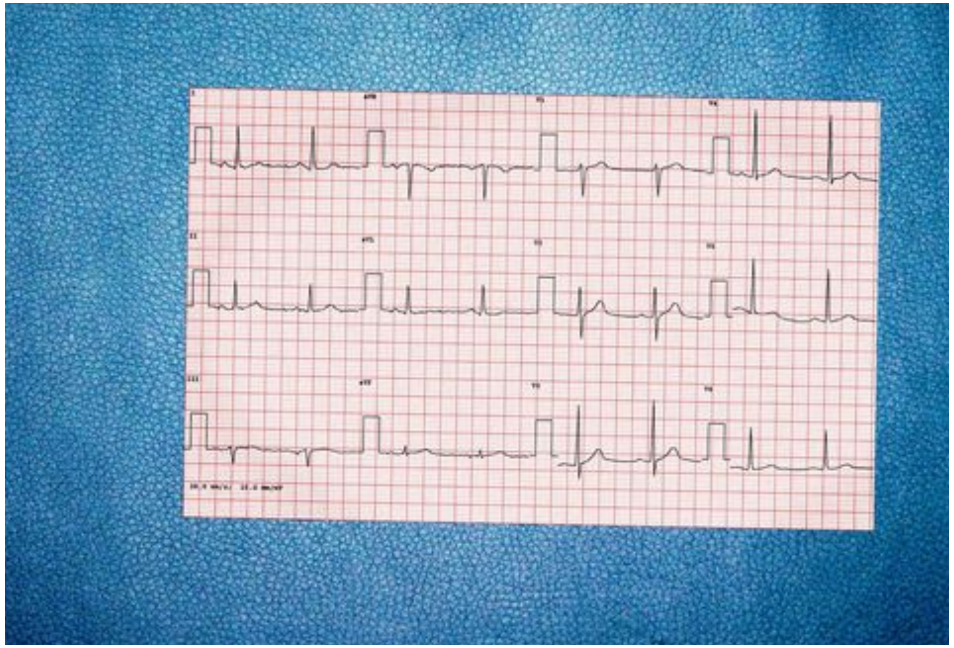
3 ECG Processing Overview

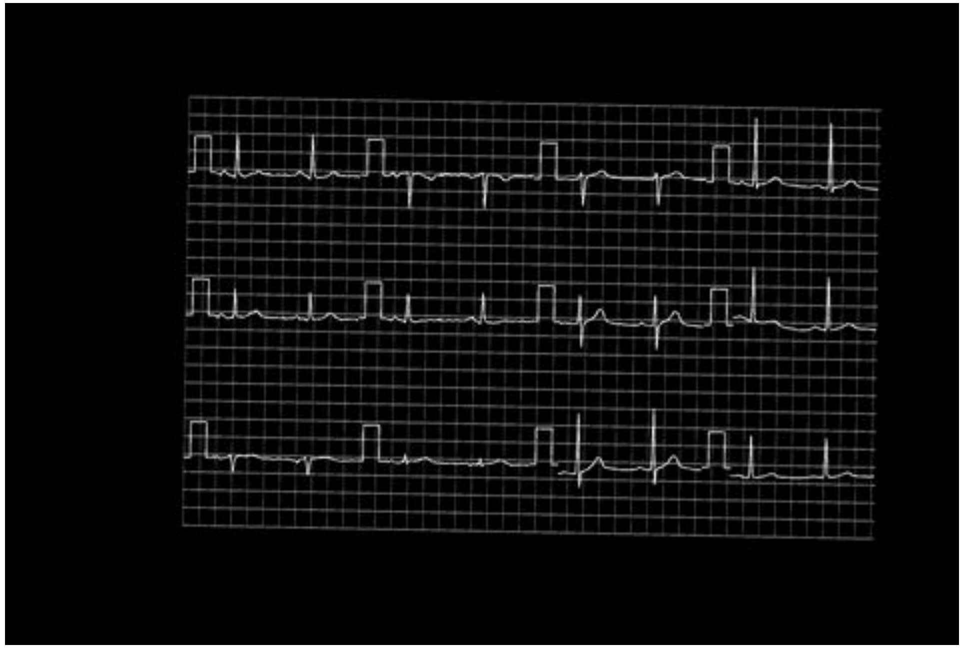


Ronneberger, O., Fischer, P., & Brox, T.. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. : https://arxiv.org/abs/1505.04597

3.1 ECG Segmentation

The U-Net was trained with pairs of masks and generated ECG diagrams with data augmentation applied to them, notice that the mask corresponds to the training label.





Grid-lines are also captured in the training to be used for auto-orienting of diagrams during processing.

3.1 ECG Segmentation

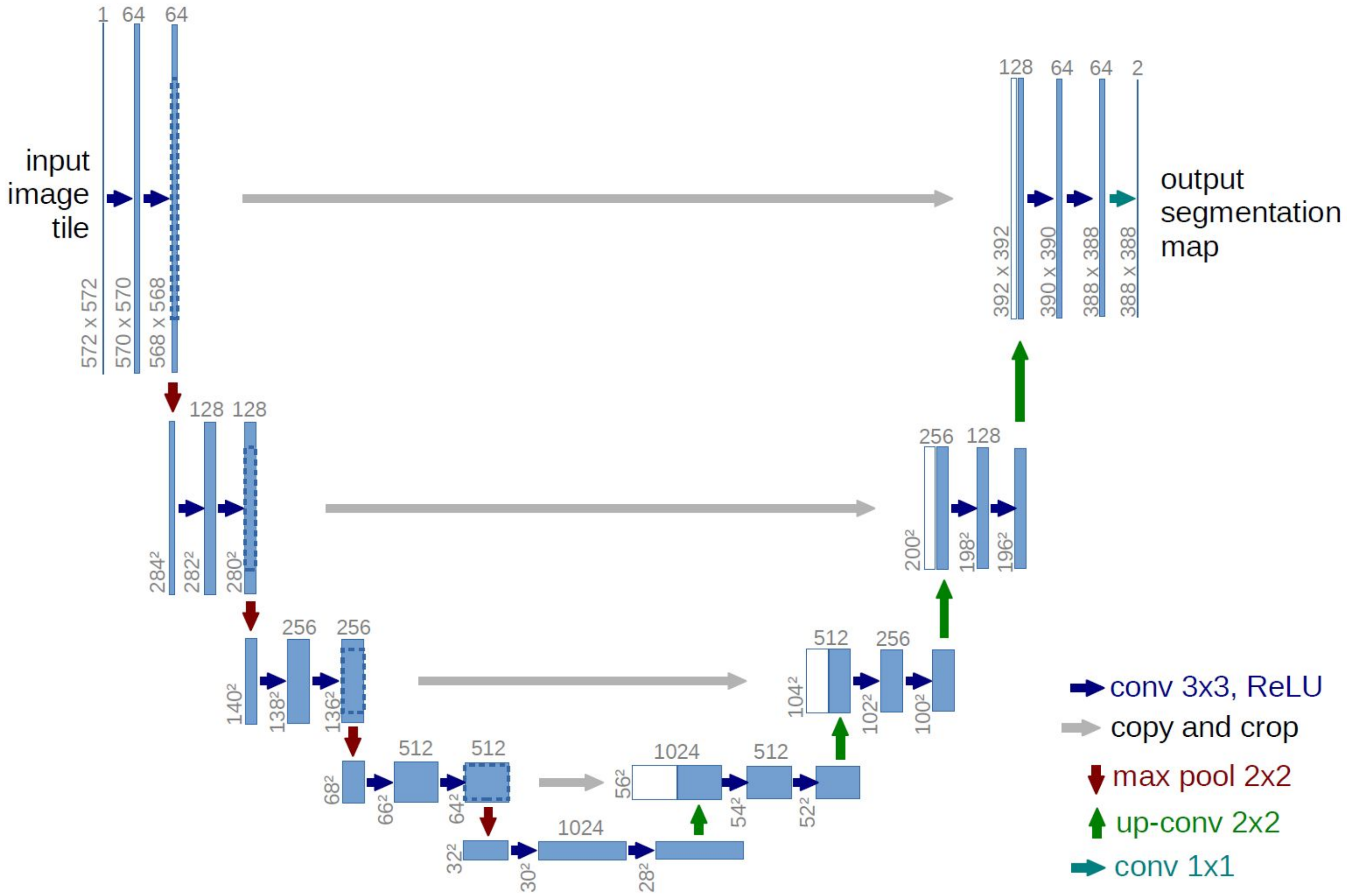
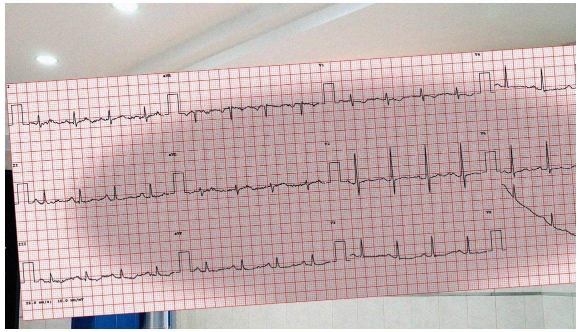
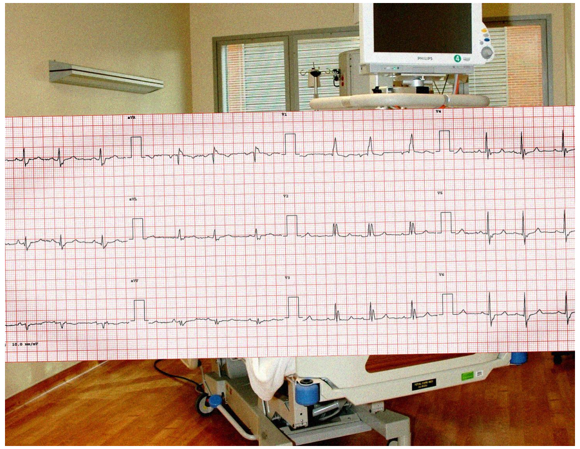
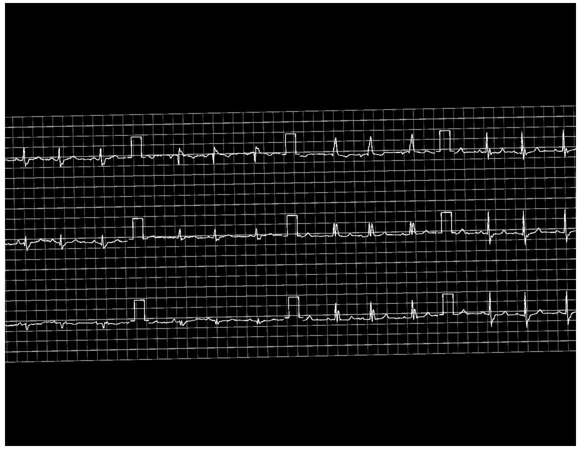


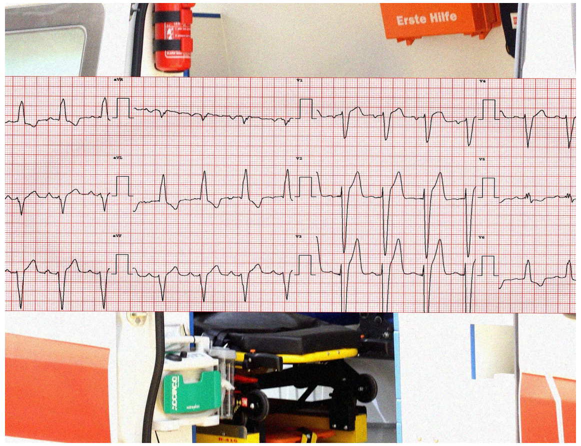
Figure taken from: Ronneberger, O., Fischer, P., & Brox, T.. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. : https://arxiv.org/abs/1505.04597

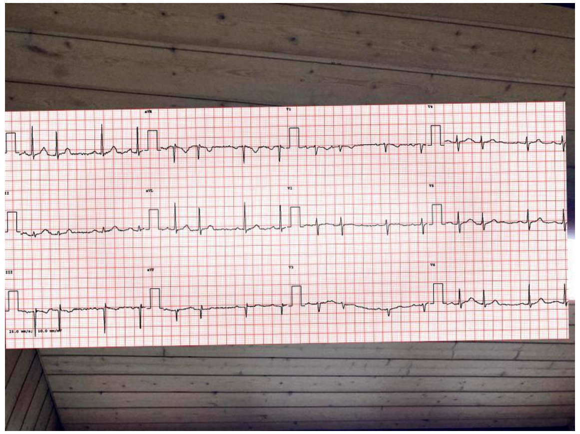
3.1 ECG Segmentation

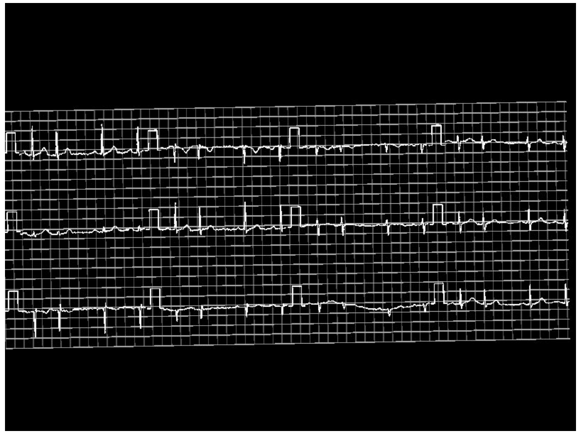




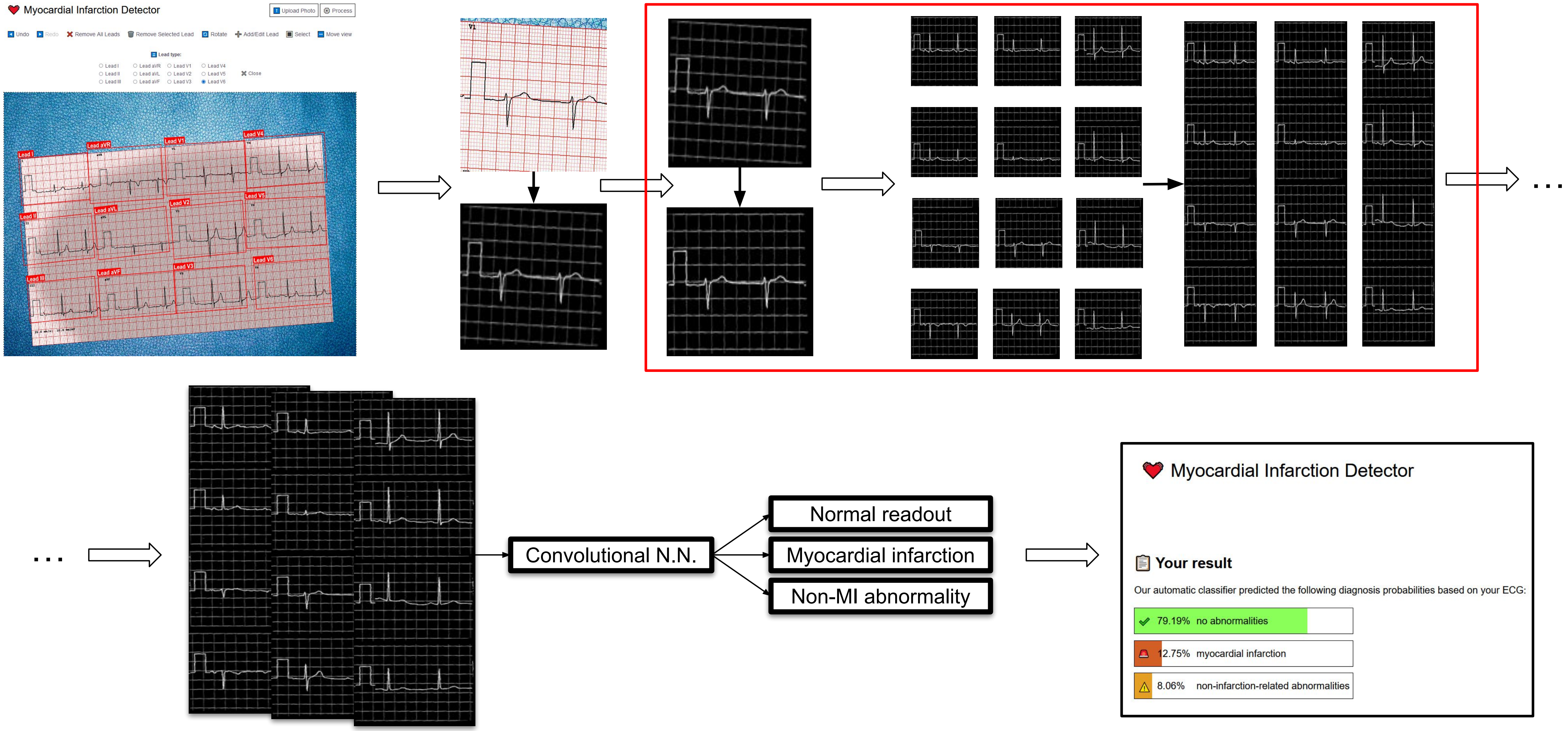




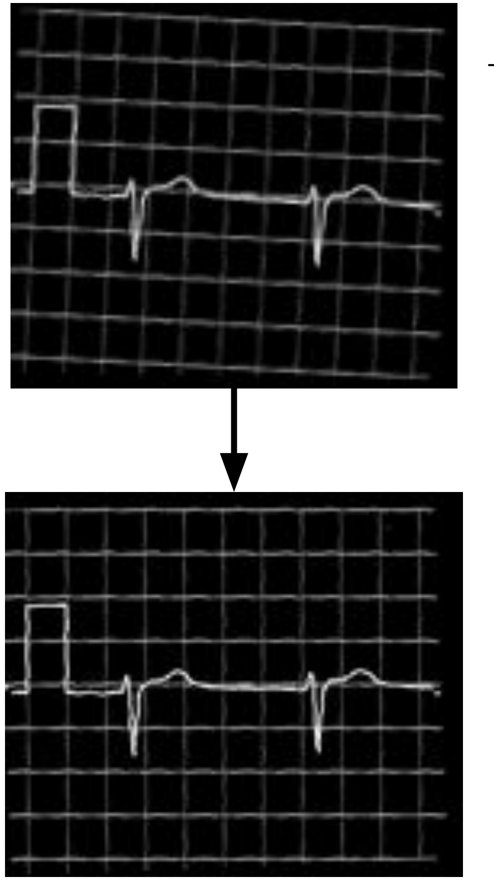


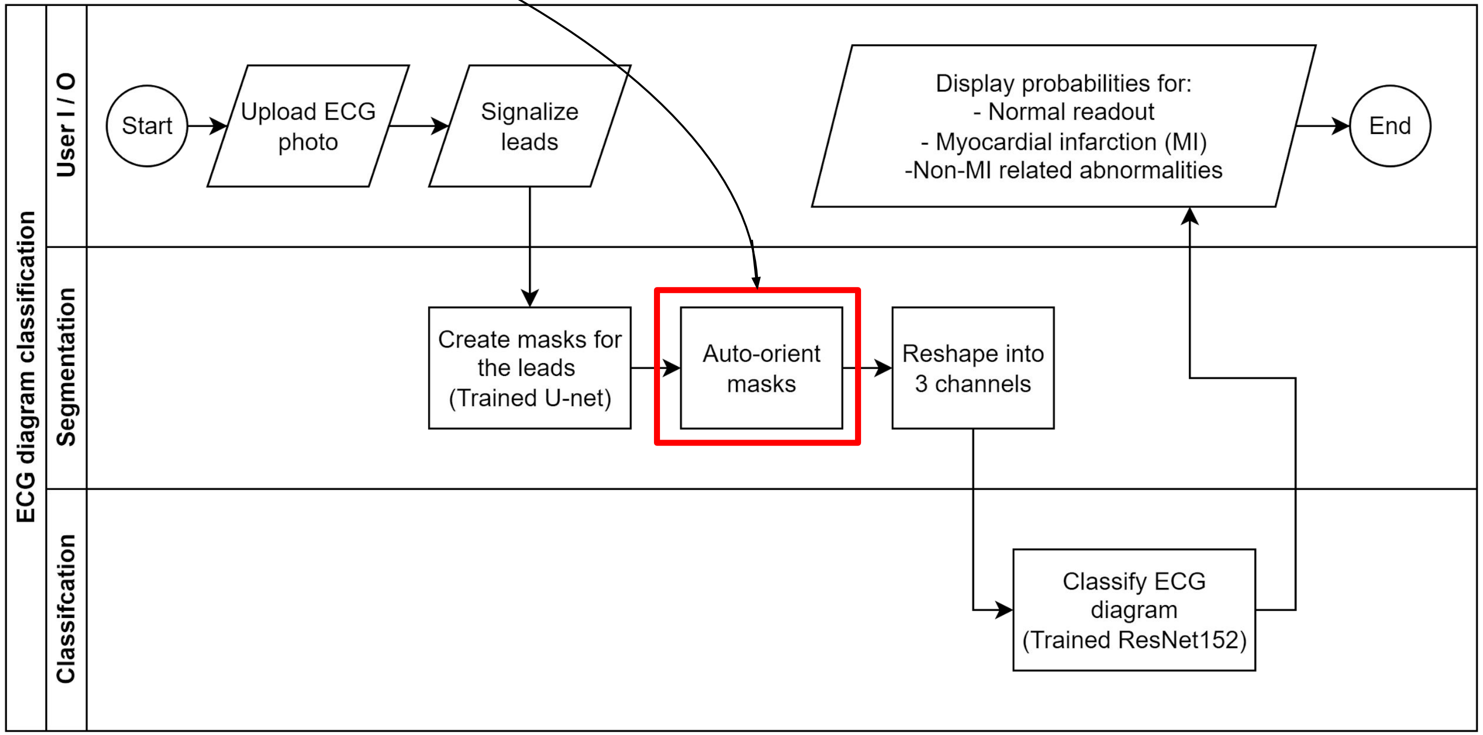


3 ECG Processing Overview

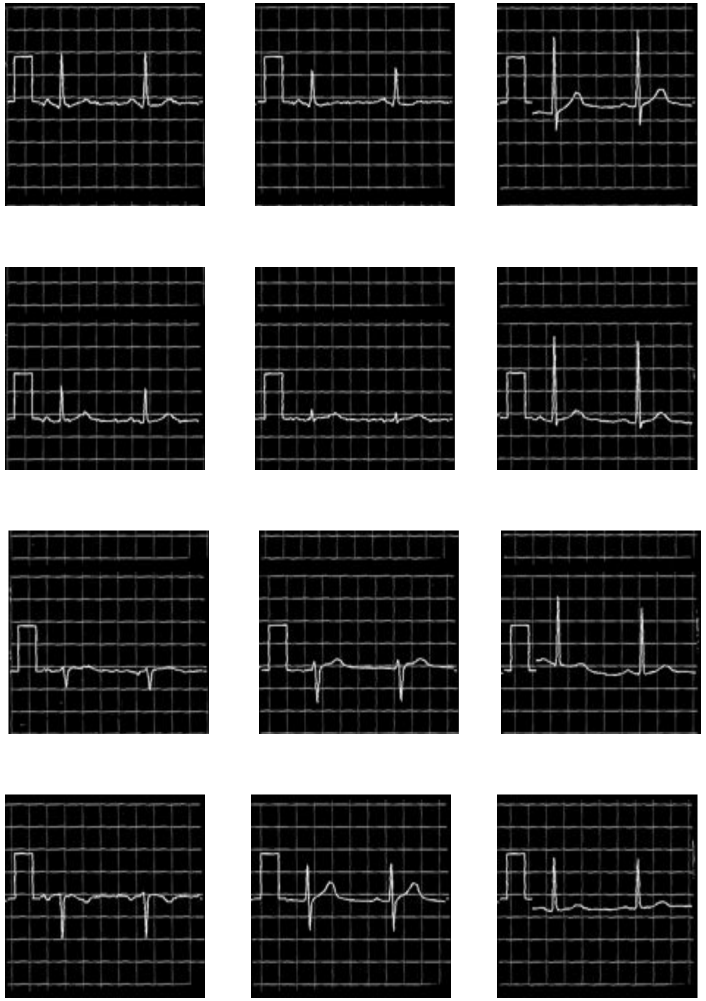


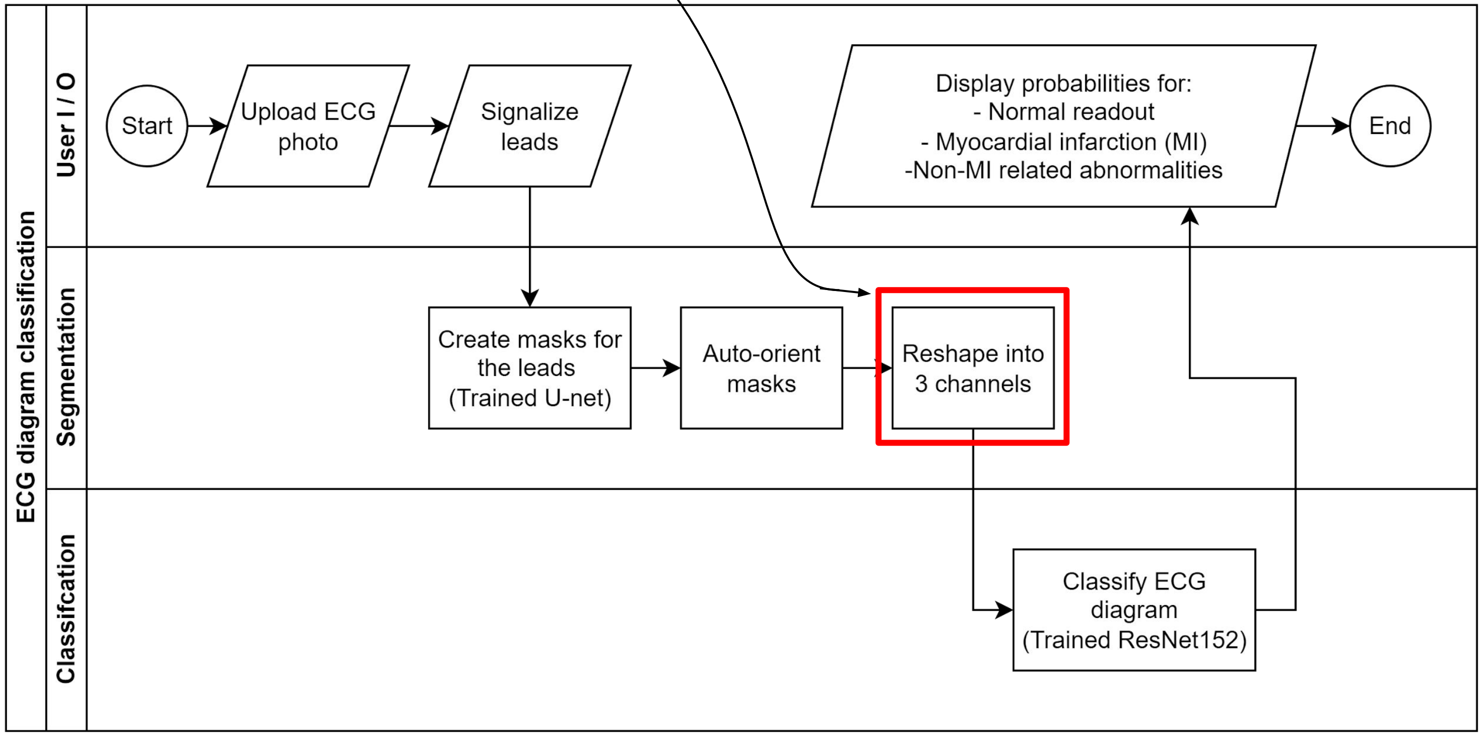
3 ECG Processing Overview





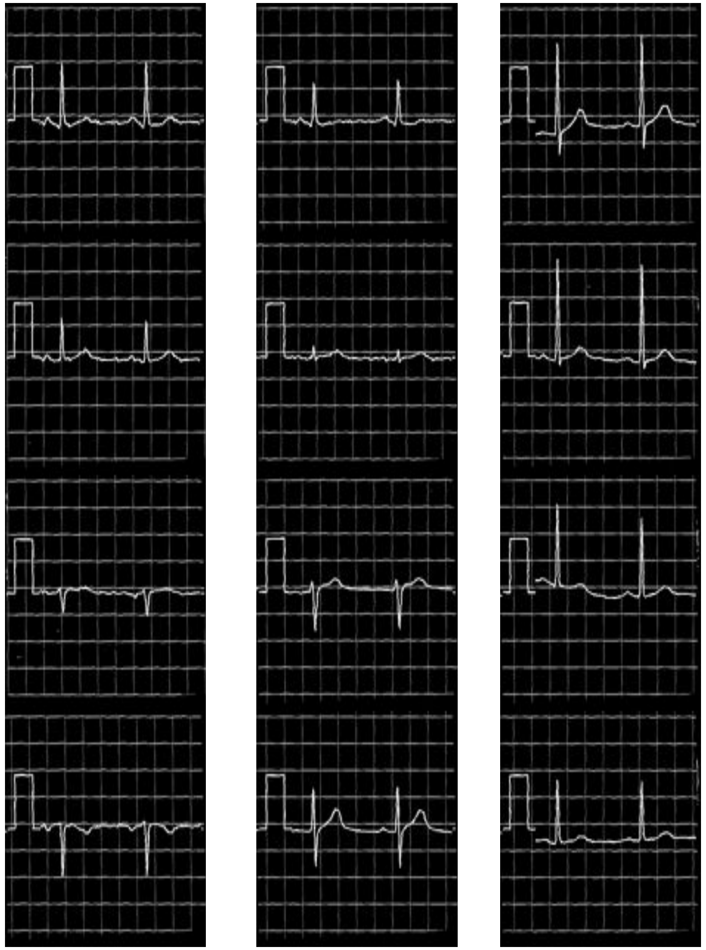
3 ECG Processing Overview



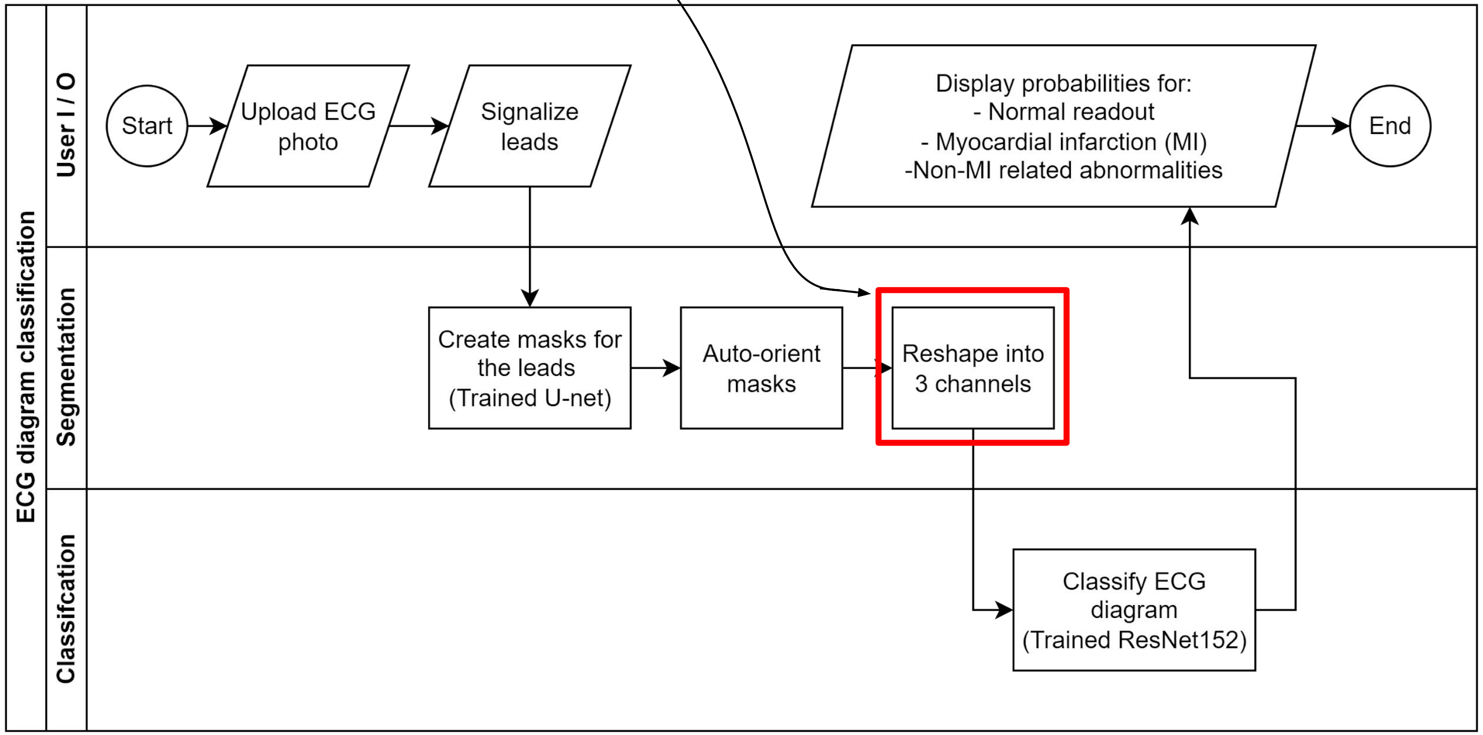


ETH zürich

3 ECG Processing Overview

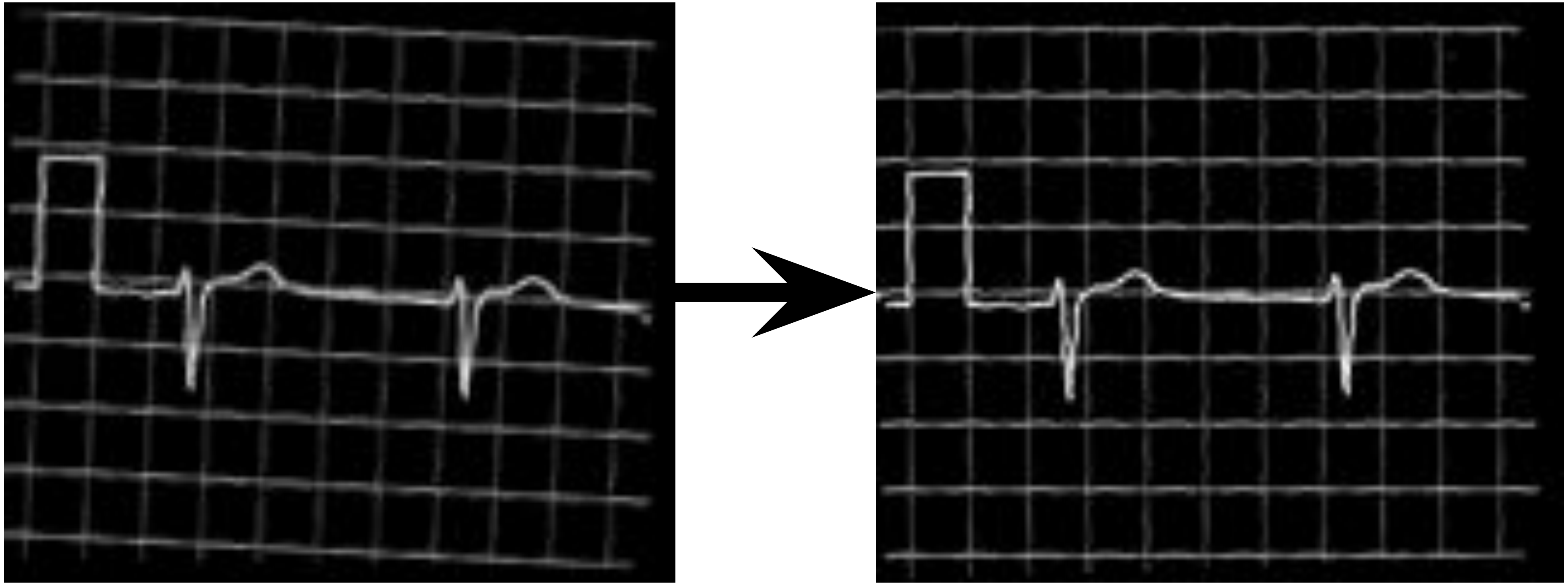


Channel 1 Channel 2 Channel 3



3.1 ECG Segmentation

The straightening algorithm is based on detecting horizontally the grid-lines on several heights and then correct the rotation

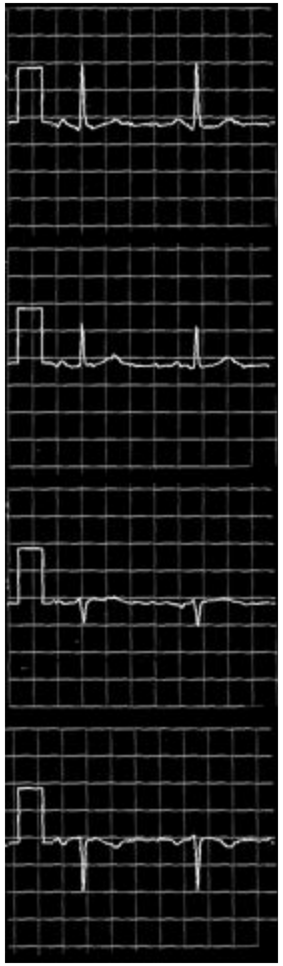


3.1 ECG Segmentation

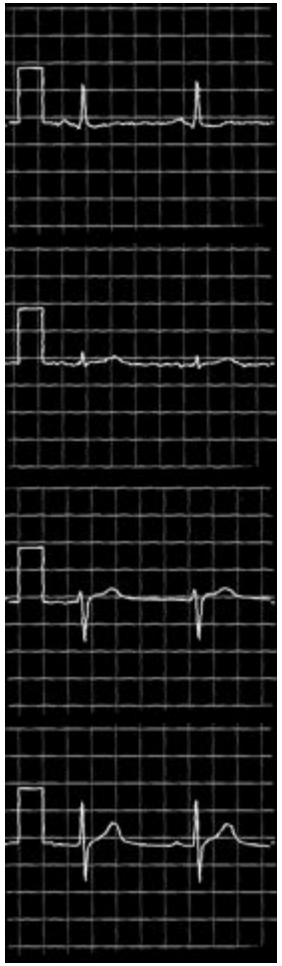
The 12 leads are divided in 3 groups, one for each channel fed into the convnet for classification.

If more leads are inserted, they are added in a **consistent** way to the different channels. **This is important because certain types of myocardial infarctions result in characteristic changes to certain leads.**

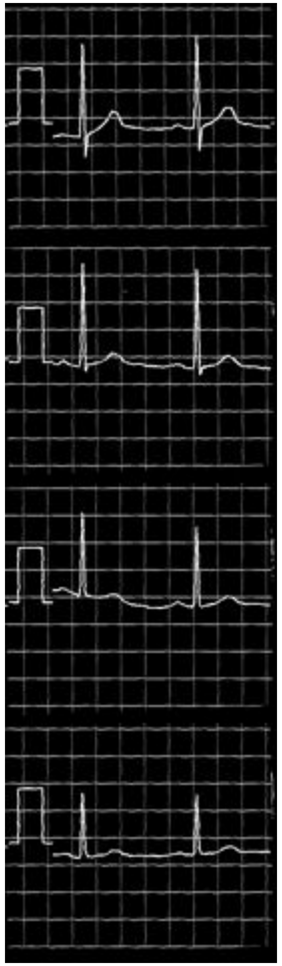
The grid lines are kept also fed to the classifying convnet. We expect that this helps with detecting correlations of different lead values at equal times.



Channel 1

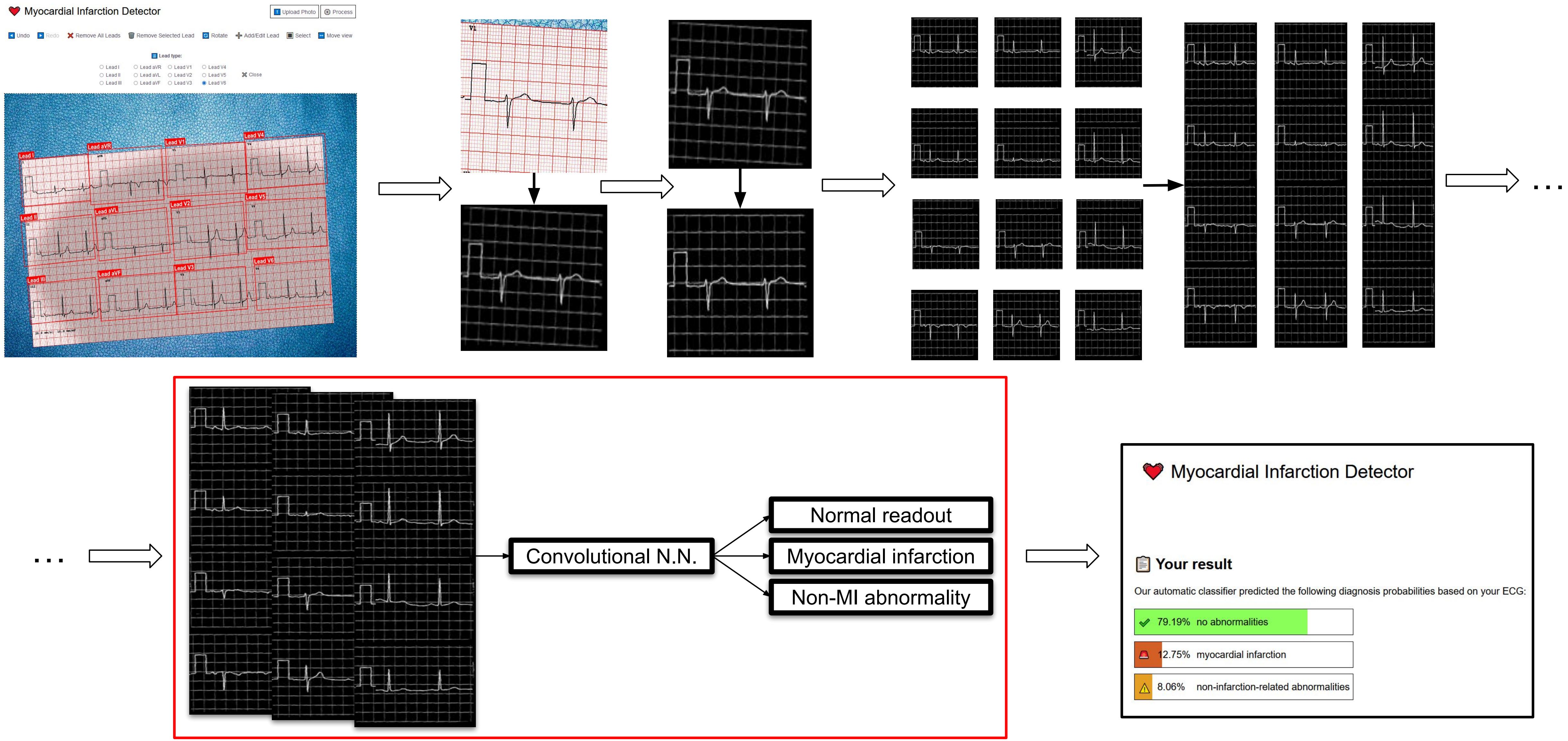


Channel 2



Channel 3

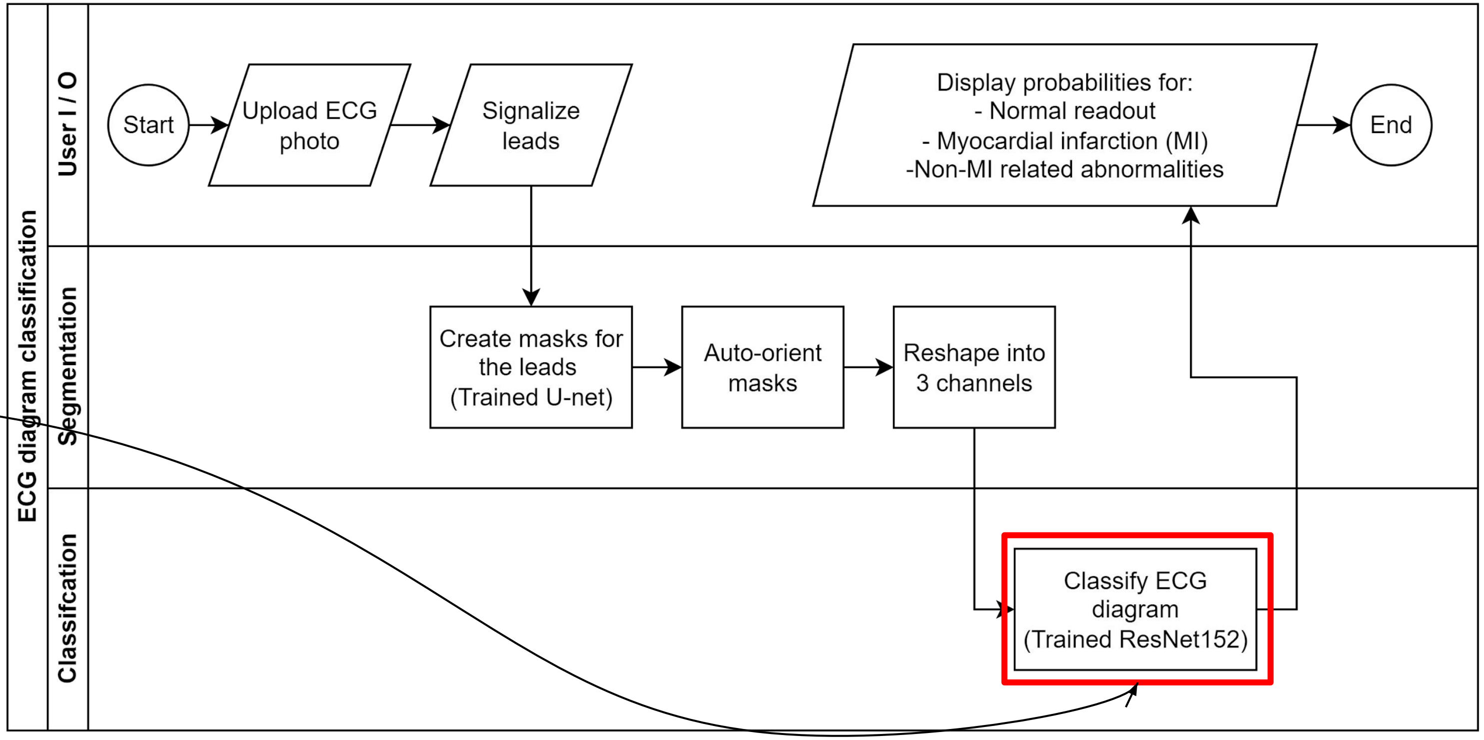
3 ECG Processing Overview



3 ECG Processing Overview

Channel 1 Channel 2 Channel 3

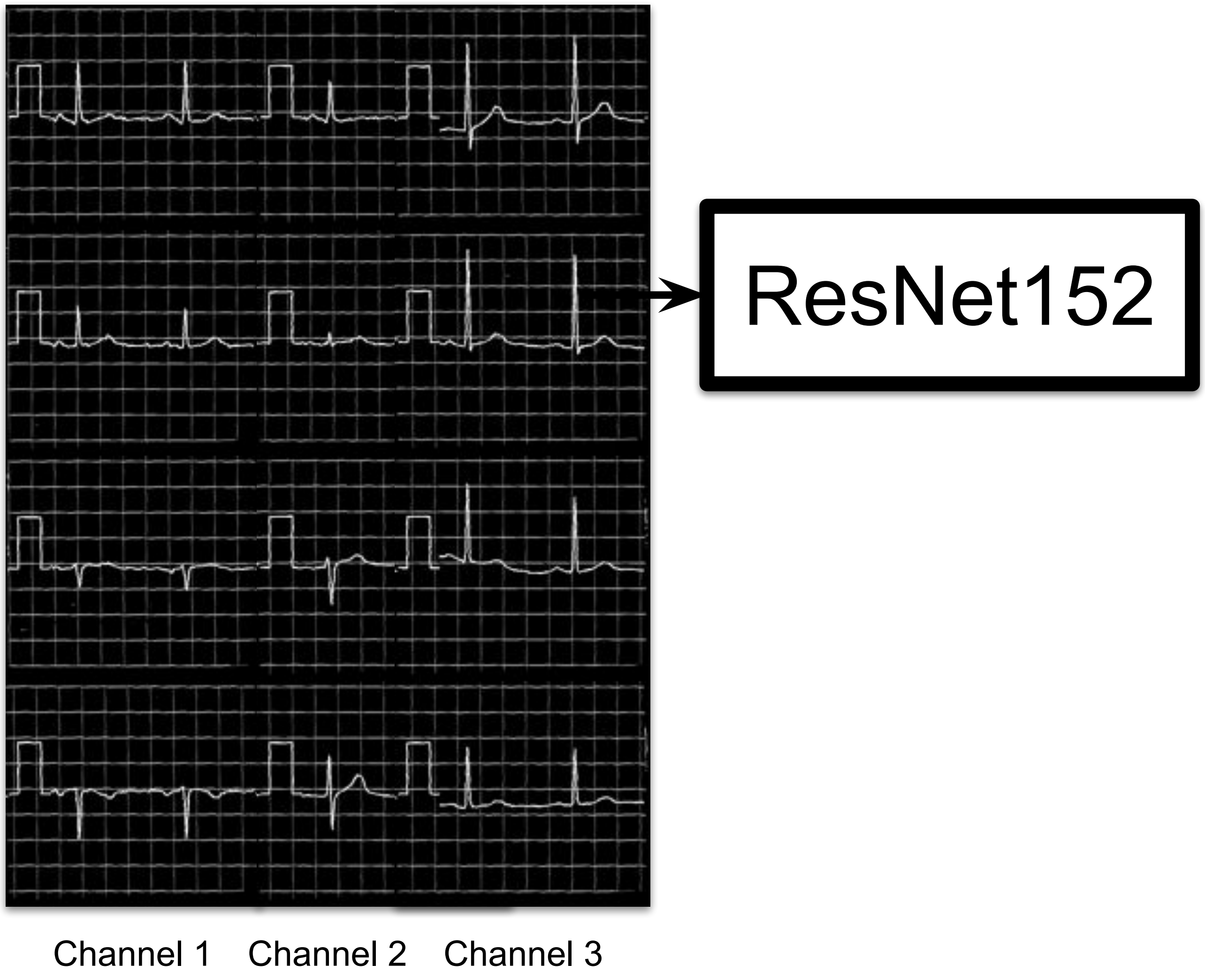
ResNet152



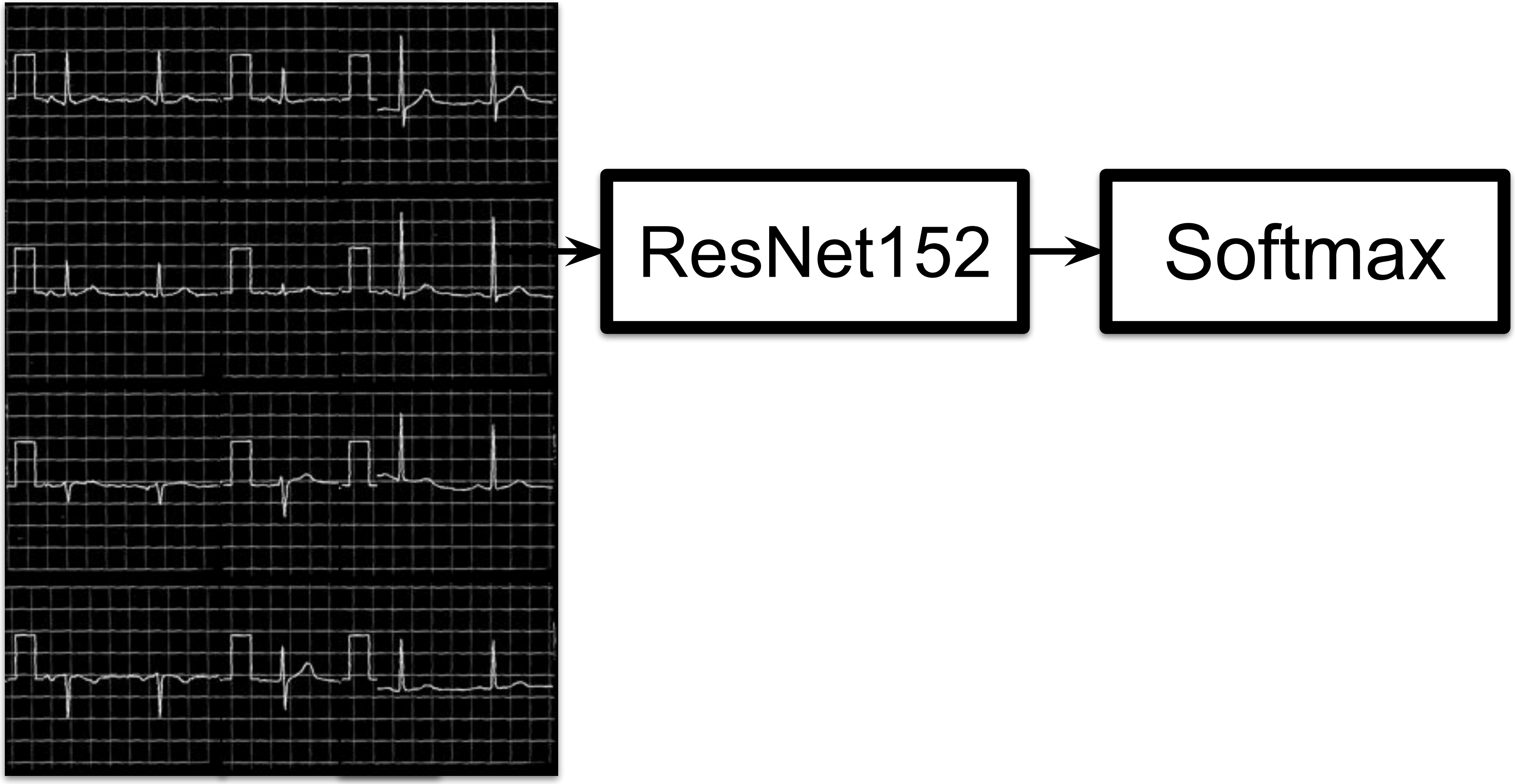
3.2 ECG Classification

Channel 1 Channel 2 Channel 3

3.2 ECG Classification



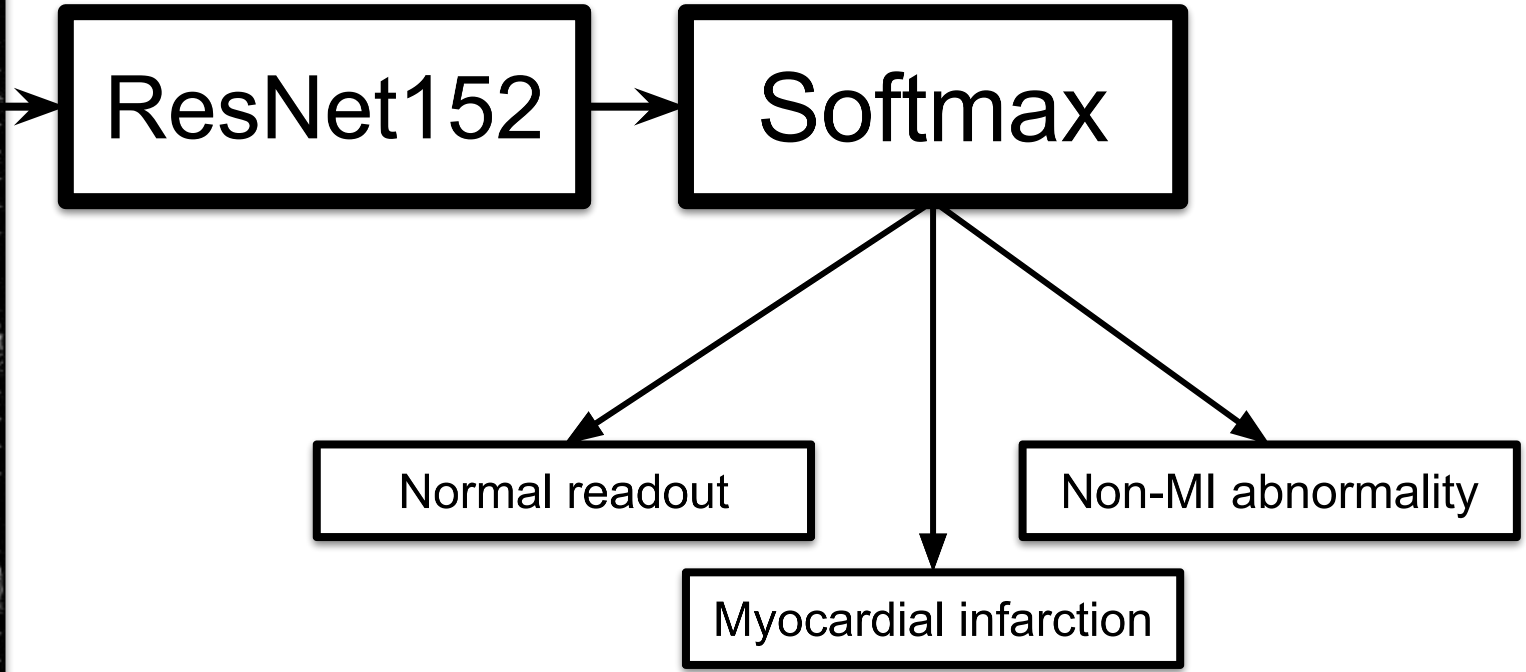
3.2 ECG Classification



Channel 1 Channel 2 Channel 3

3.2 ECG Classification

Channel 1 Channel 2 Channel 3



3.2 ECG Classification: ResNet152 architecture

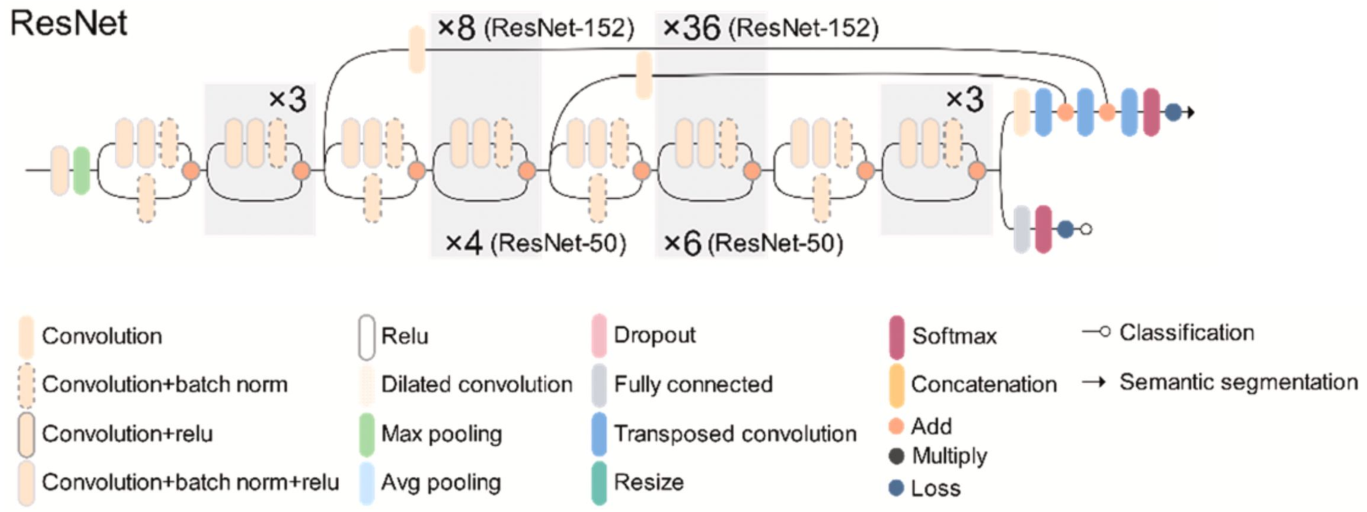


Figure 1: ResNet152 architecture diagram

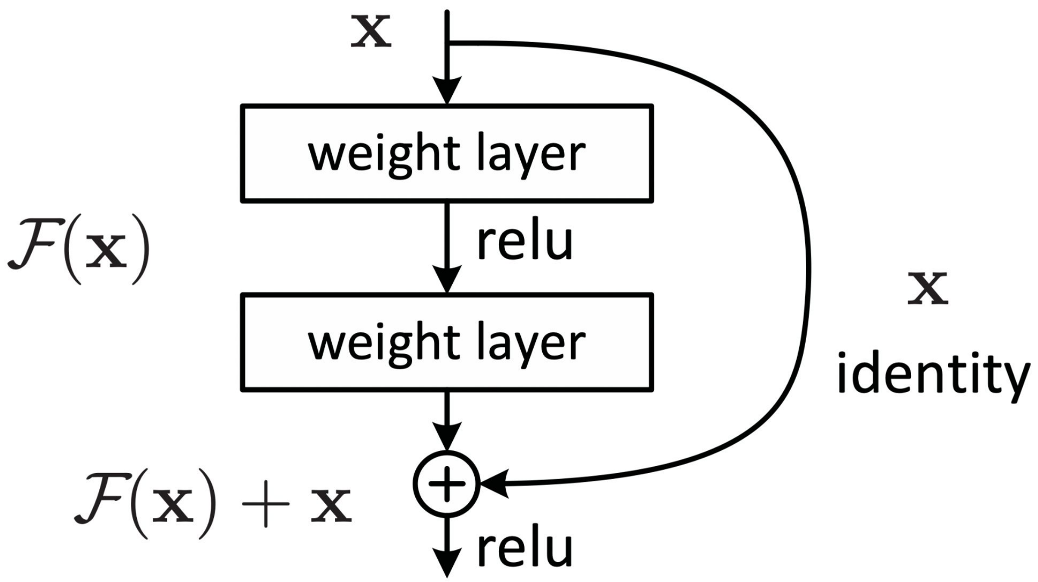
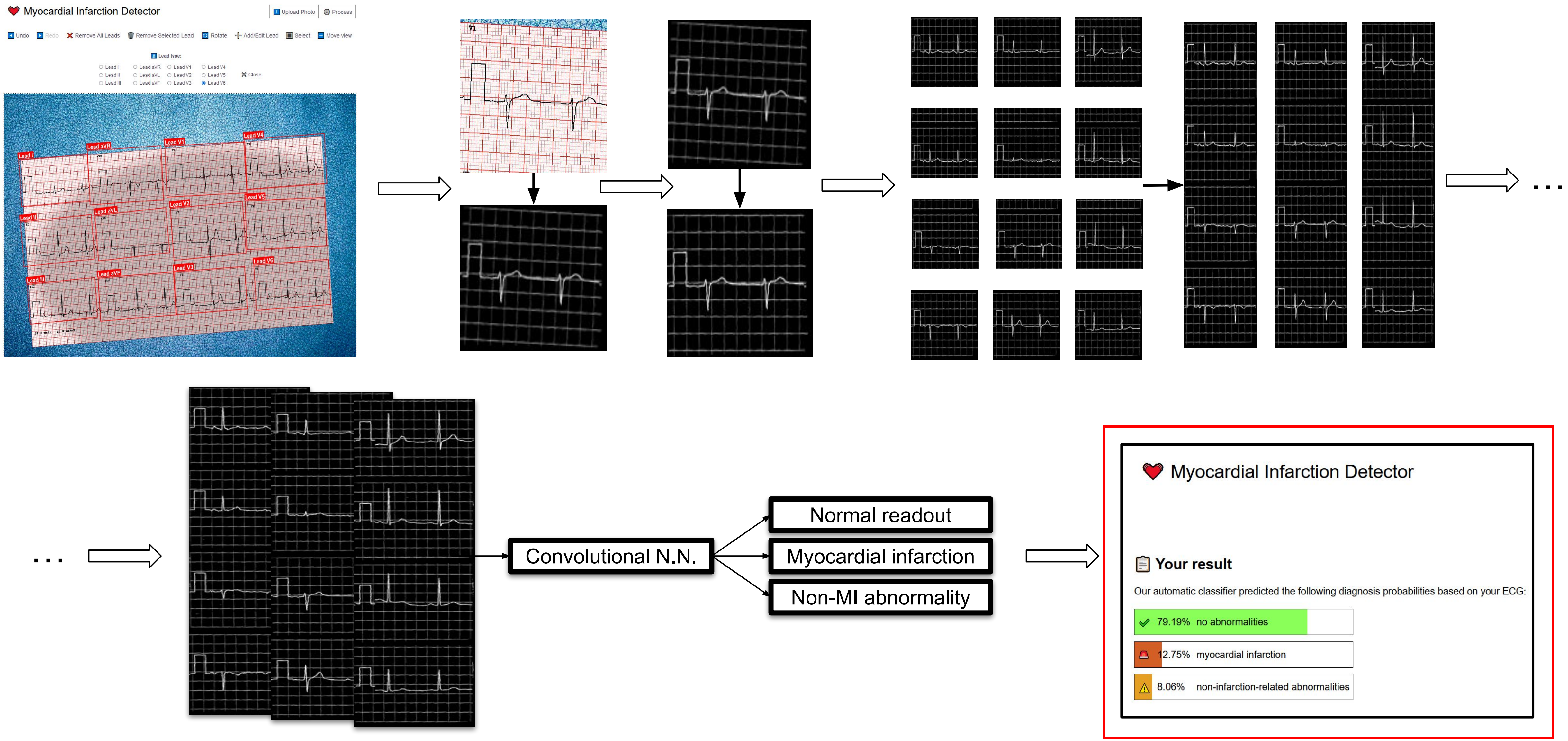


Figure 2: Residual learning building block

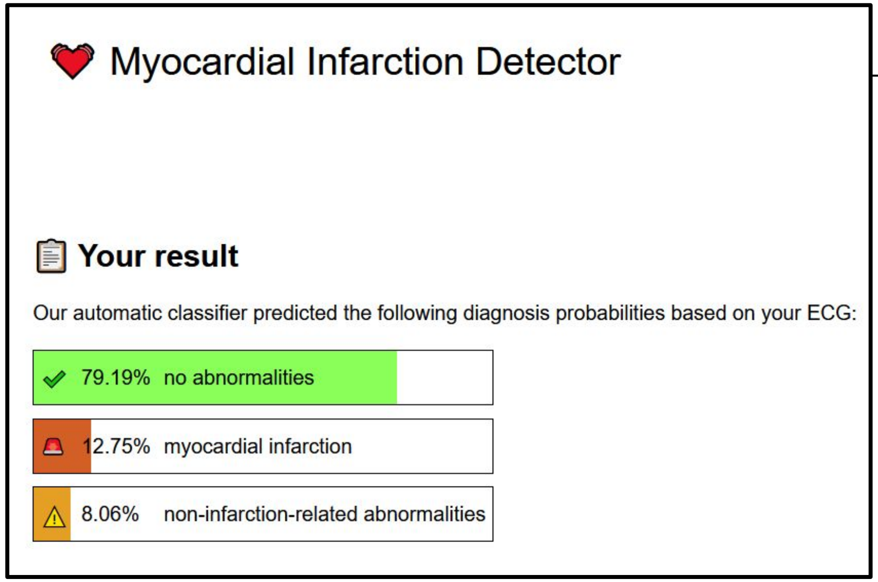
Figure 1 taken from: Zhouxin Xi, Chris Hopkinson, Stewart B. Rood, & Derek R. Peddle (2020). See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing, 168, 1-16.

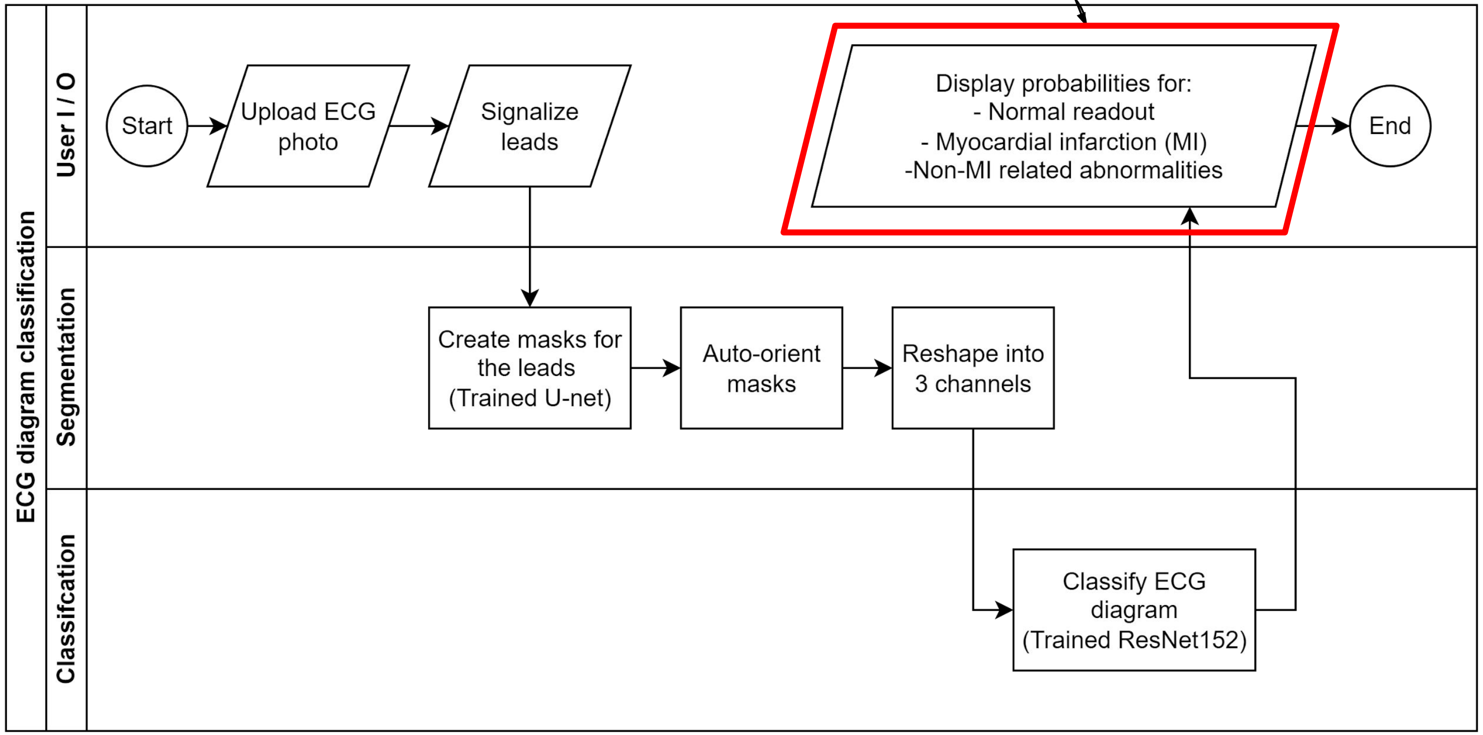
Figure 2 taken from: Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun (2015). Deep Residual Learning for Image Recognition. CoRR, https://arxiv.org/abs/1512.03385v1

3 ECG Processing Overview



3 ECG Processing Overview

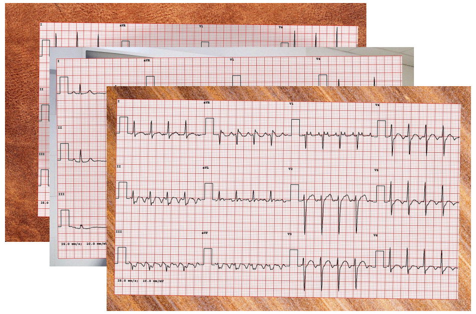


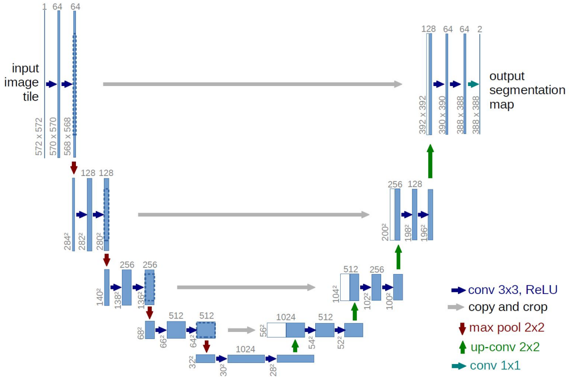


This will be presented with detail later on.

3.3 Classification Network Training

Extraction of segmentation images

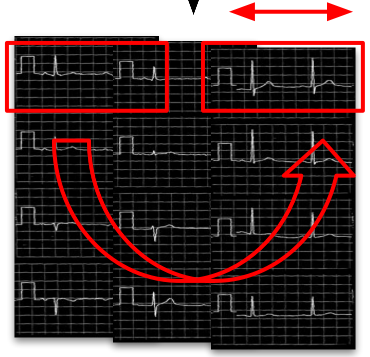




<u>dataset stats:</u>

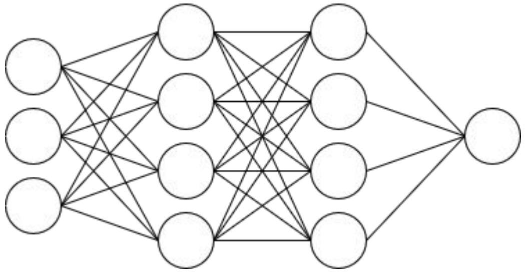
<u>before augmentation: 13826 ECGs</u>
6152 normal ECGs
4752 ECGs with non-MI abnormalities
2922 ECGs showing MIs

<u>after augmentation: 41478 ECGs</u>
18456 normal ECGs
14256 ECGs with non-MI abnormalities
8766 ECGs showing MIs



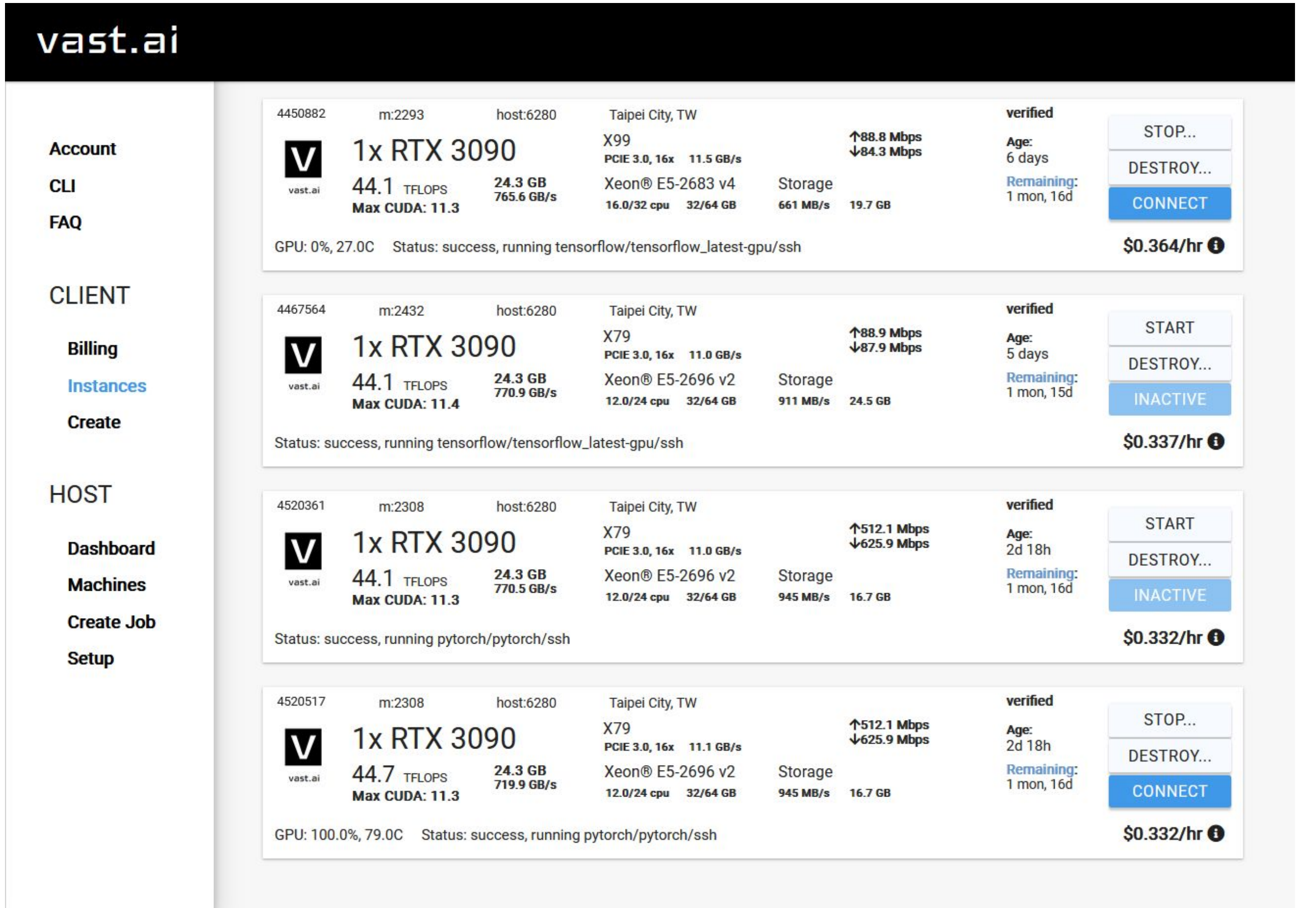
Swapping and stretching of leads

Data augmentation



Training on augmented segmentation images

3.3 Classification & Segmentation Network Training: Training Environment



Remote development & training using vast.ai

3.3 Classification & Segmentation Network Training: Training Frameworks

Frameworks:

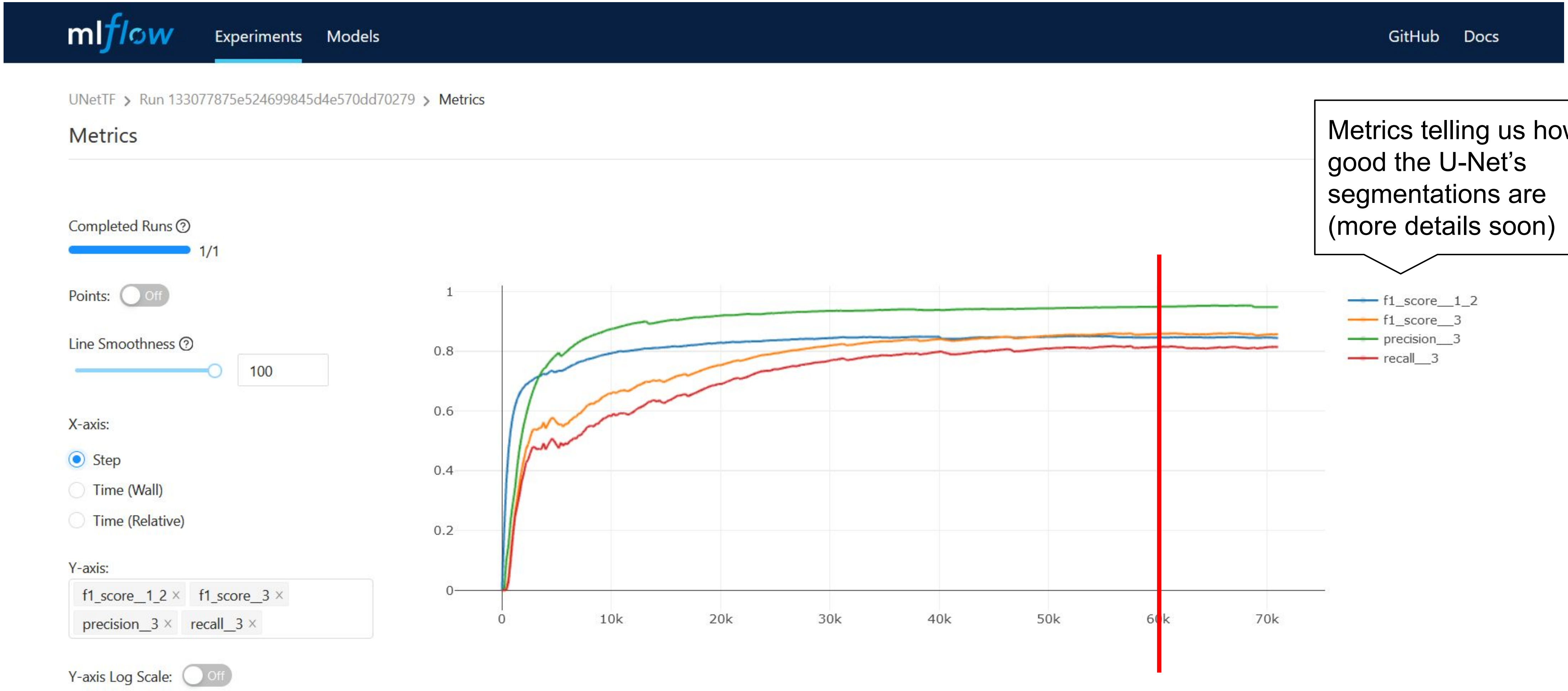
fast.ai

mlflow™

TensorFlow

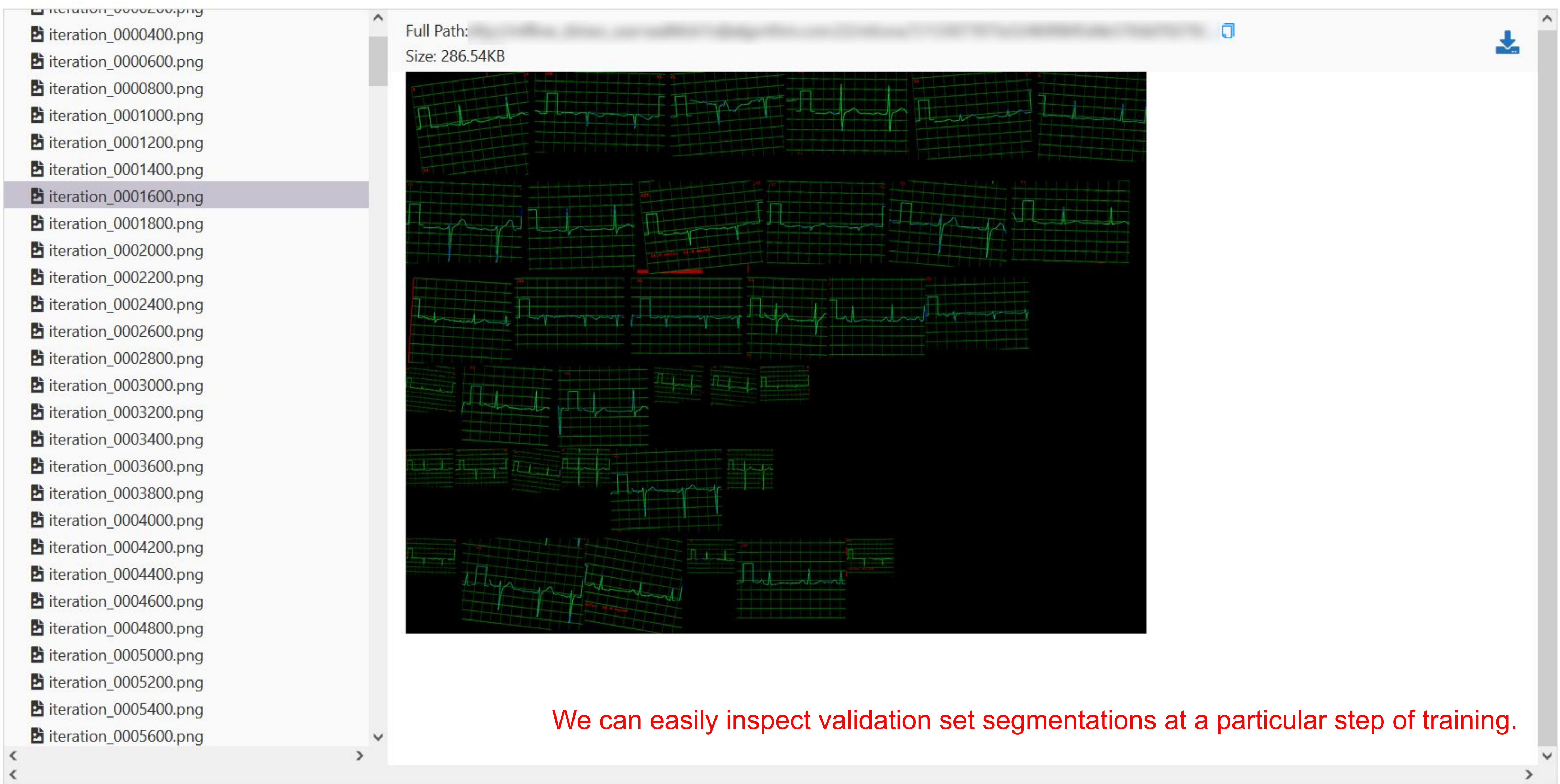
For classification network

For segmentation network



Metrics telling us how good the U-Net's segmentations are (more details soon)

We can easily see (approximate) the step at which convergence is achieved.

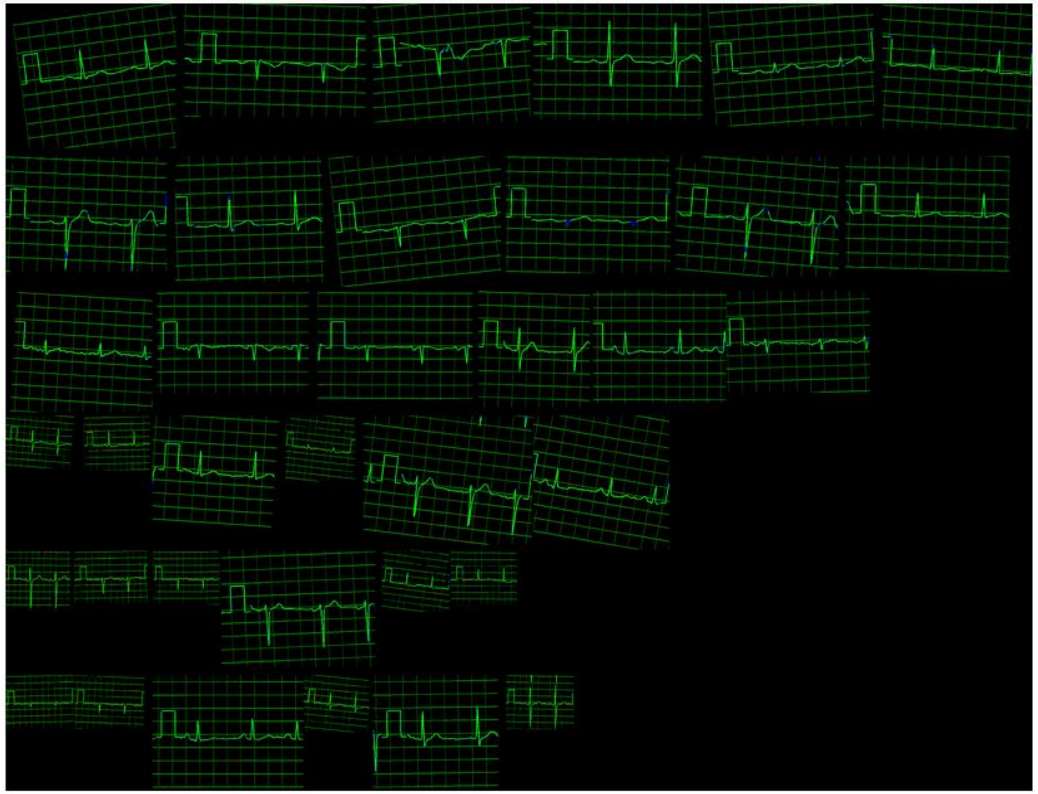


We can easily inspect validation set segmentations at a particular step of training.

iteration_0058200.png
iteration_0058400.png
iteration_0058600.png
iteration_0058800.png
iteration_0059000.png
iteration_0059200.png
iteration_0059400.png
iteration_0059600.png
iteration_0059800.png
iteration_0060000.png
iteration_0060200.png
iteration_0060400.png
iteration_0060600.png
iteration_0060800.png
iteration_0061000.png
iteration_0061200.png
iteration_0061400.png
iteration_0061600.png
iteration_0061800.png
iteration_0062000.png
iteration_0062200.png
iteration_0062400.png
iteration_0062600.png
iteration_0062800.png
iteration_0063000.png
iteration_0063200.png
iteration_0063400.png
iteration_0063600.png

Full Path: [redacted]

Size: 249.8KB



…after convergence, visualizations look as they should (no red spots ➔ no errors)

3.3 Classification Network Training: Metrics

Precision:

$$\frac{\#\ \text{True positives}}{\#\ \text{True positives} + \#\ \text{False positives}} \quad \in [0, 1]$$

i.e. "if the classifier predicts an MI, how likely is it really an MI?"

Recall:

- - -

F1 score:

- - -

3.3 Classification Network Training: Metrics

Precision:

$$\frac{\#\ \text{True positives}}{\#\ \text{True positives} + \#\ \text{False positives}} \in [0, 1]$$

i.e. “if the classifier predicts an MI, how likely is it really an MI?”

Recall:

$$\frac{\#\ \text{True positives}}{\#\ \text{True positives} + \#\ \text{False negatives}} \in [0, 1]$$

i.e. “if there is an MI, how likely is the classifier to say so?”

F1 score:

\- - -

3.3 Classification Network Training: Metrics

Precision:

$$\frac{\#\ \text{True positives}}{\#\ \text{True positives} + \#\ \text{False positives}} \quad \in [0, 1]$$

i.e. “if the classifier predicts an MI, how likely is it really an MI?”

Recall:

$$\frac{\#\ \text{True positives}}{\#\ \text{True positives} + \#\ \text{False negatives}} \quad \in [0, 1]$$

i.e. “if there is an MI, how likely is the classifier to say so?”

F1 score:

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \in [0, 1]$$

harmonic mean of precision and recall; means to combine the two scores

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	- - -	- - -	- - -
Myocardial Infarction	- - -	- - -	- - -

Results on validation dataset of augmented ECG segmentation images

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	- - -	- - -	- - -

Results on validation dataset of augmented ECG segmentation images

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	**52.81%**	**66.00%**	**0.5868**

Results on validation dataset of augmented ECG segmentation images

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	**52.81%**	**66.00%**	**0.5868**

Results on validation dataset of augmented ECG segmentation images

validation dataset:
1331 ECG segmentations

581 normal ECGs (43.7%)

462 ECGs w/ **non-MI abnormalities** (34.7%)

288 ECGs w/ **MI** (21.6%)

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	**52.81%**	**66.00%**	**0.5868**

Results on validation dataset of augmented ECG segmentation images

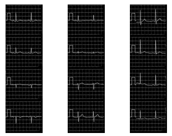
validation dataset:
1331 ECG segmentations

581 normal ECGs (43.7%)

462 ECGs w/ **non-MI abnormalities** (34.7%)

288 ECGs w/ **MI** (21.6%)

- Big problem: **classification networks overfit!**
  - Cannot perform data augmentation on ECG segmentation images easily on-the-fly
    - Need to **synthesize new ECG photographs**
    - **Time-consuming** process!



3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	**52.81%**	**66.00%**	**0.5868**

Results on validation dataset of augmented ECG segmentation images

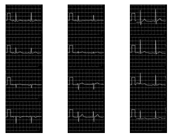
validation dataset:
1331 ECG segmentations

581 normal ECGs (43.7%)

462 ECGs w/ **non-MI abnormalities** (34.7%)

288 ECGs w/ **MI** (21.6%)

- ➢ Big problem: **classification networks overfit!**
  - ➢ Cannot perform data augmentation on ECG segmentation images easily on-the-fly
    - ➢ Need to **synthesize new ECG photographs**
    - ➢ **Time-consuming** process!



- ➢ One study (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8273385/) using **10-second time series data** (raw PTB-XL dataset) achieved **precision of 78.9%**, **recall of 81.8%**, **F1 score of 0.825 on MIs (*binary* classification)**

3.3 Classification Network Training: Results

Class	Precision	Recall	F1 Score
Normal ECG	73.17%	88.87%	0.8027
Non-MI-related abnormalities	71.55%	41.09%	0.5220
Myocardial Infarction	**52.81%**	**66.00%**	**0.5868**

Results on validation dataset of augmented ECG segmentation images

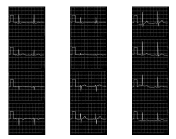
validation dataset:
1331 ECG segmentations

581 normal ECGs (43.7%)

462 ECGs w/ **non-MI abnormalities** (34.7%)

288 ECGs w/ **MI** (21.6%)

- ➢ Big problem: **classification networks overfit!**
  - ➢ Cannot perform data augmentation on ECG segmentation images easily on-the-fly



    - ➢ Need to **synthesize new ECG photographs**
    - ➢ **Time-consuming** process!
- ➢ One study (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8273385/) using **10-second time series data** (raw PTB-XL dataset) achieved **precision of 78.9%**, **recall of 81.8%**, **F1 score of 0.825 on MIs (*binary* classification)**
- ➢ We view our **result** as a **reasonable proof-of-concept**, considering we **only take snippets of ~3 seconds** for each lead and **use *distorted image data*** rather than pure time series, and did not perform hyperparameter tuning
  - ➢ **Shorter lead recordings** are much more **common in practice**

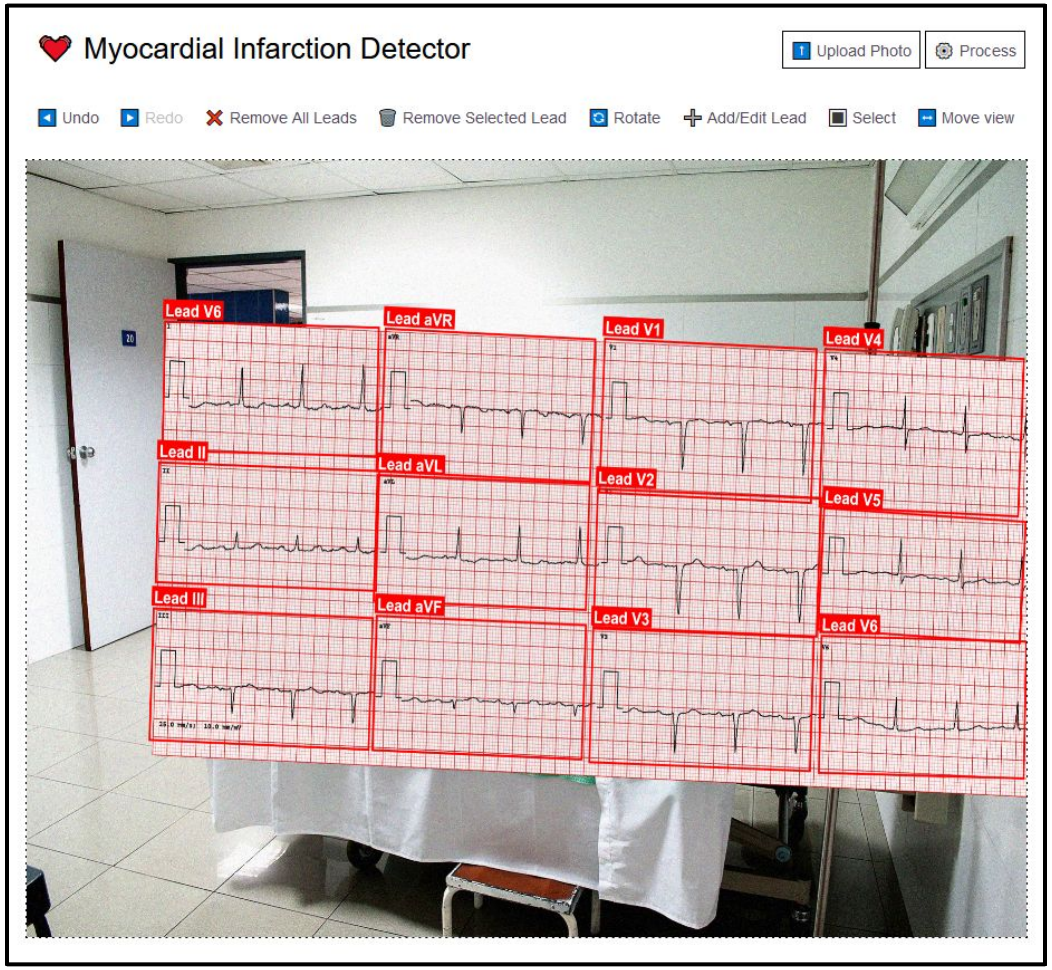
3.3 Classification Network Training: Precision-Recall-F1 Table For MI Class

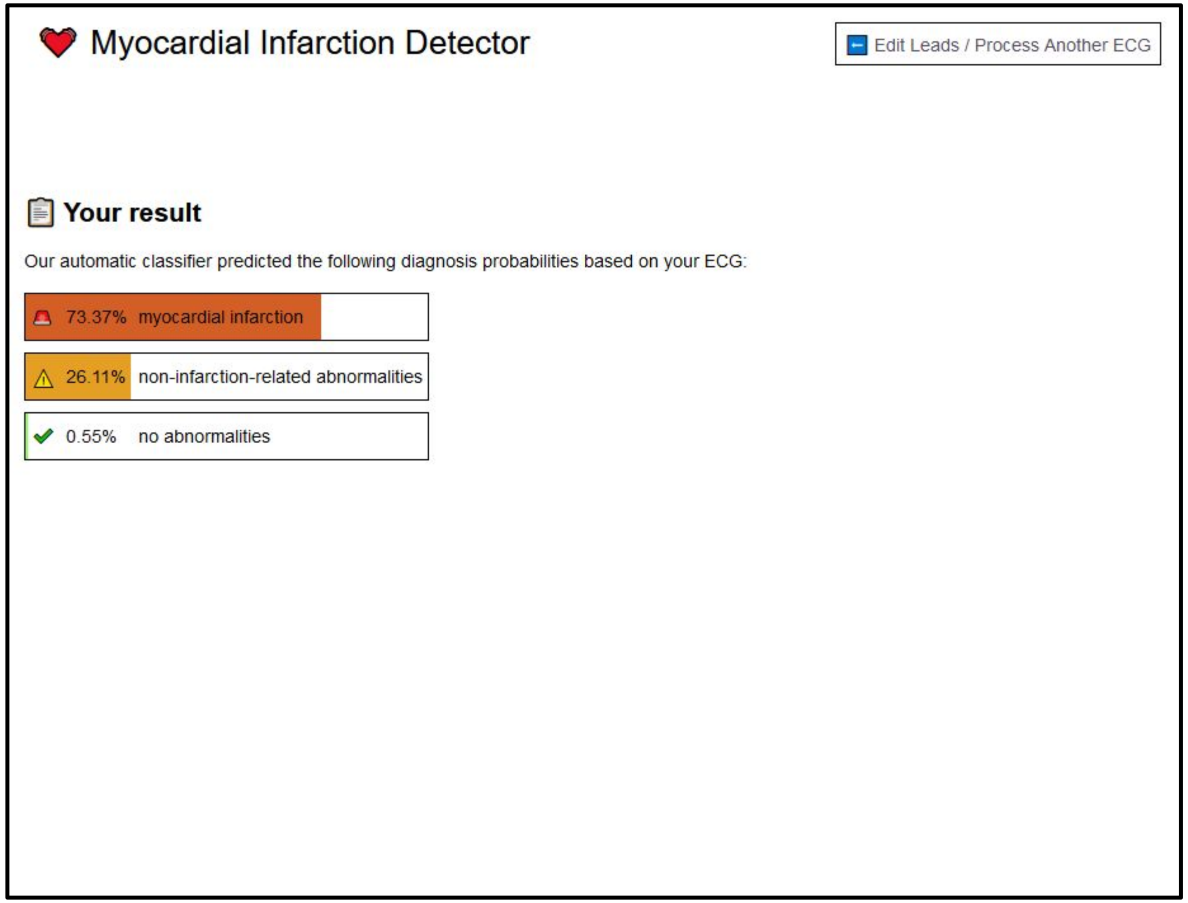
Precision	Recall	F1 Score
68.93%	49.31%	0.5749
67.44%	50.69%	0.5788
65.46%	52.20%	0.5808
63.61%	54.17%	0.5851
62.14%	55.97%	0.5890
60.81%	57.64%	0.5918
59.31%	59.08%	0.5919
58.02%	60.46%	0.5921

Precision	Recall	F1 Score
56.66%	61.84%	0.5914
55.38%	63.60%	0.5907
54.16%	64.74%	0.5898
52.81%	**66.00%**	**0.5868**
51.45%	67.28%	0.5831
50.11%	68.58%	0.5791
48.75%	69.79%	0.5741
47.42%	70.94%	56.85%

➔ add artificial bias to classifier’s score for MI class (start: +0; Δ: +0.0263 with each row)
➔ determine optimal / good boost: **should prioritize recall over precision!**
➔ how to trade off? cost of human life (false negative) vs. wasted time/wrong treatment (false positive)

3.4 Web Tool





Current address (temporary): http://algvrithm.com:8080/

4. Future Work

- Use even more data!
  - Can generate nearly unlimited data, but: computationally expensive
  - Move image generation logic to GPU
- Diversify training set
  - Use different lead layouts & lengths
  - Use only subset of 12 leads
- Add support for paper speeds other than default 25mm/s
- Try more architectures, perform hyperparameter tuning for segmentation and classification
- Try to detect lead locations automatically, or provide initial suggestion for user to correct
- Implement noise removal algorithms
  - e.g. correction of baseline drift
- **Test using real ECGs!**

DEMO

ETH zürich

THANK YOU!

💓

Q&A

ETH zürich

Backup slides

3.3 ECG Network Training: Training Frameworks