Supplementary material

1 Implementation details

Model	Exp.	Backbone	GCN layers	EF layers	Batch size
GCN sector	sector segm.	MobileNet2	[4,8,8,16,16,32,32,48]	-	128
NLE EF GCN	vol from kpts	r3d_18	[16,32,32,48]	[16,32,32,48]	12
NLE Vol GCN	ef from kpts	r3d_18	[16,32,32,48]	[16,32,32,48]	12

Table 1. Overview of different experiment configurations. Two multi GCN configurations are listed and also a simple sector GCN model that is trained on a subset of the training data to segment the outline of the triangular EchoNet sector in the US image.

Parameter	Value
Public dataset	EchoNet
Number of images	10.030 videos excluding 90 irregular keypoint annotations
Input image size	[3,112,112,16]
Optimization	Optimizer Adam, learning rate 10^{-4}
Pre-training	Kinect-400 (Multi-frame)
Augmentations	aligned with [23] for multiple frames

Table 2. Details on data pre-training and hyperparameters

2 Additional metrics

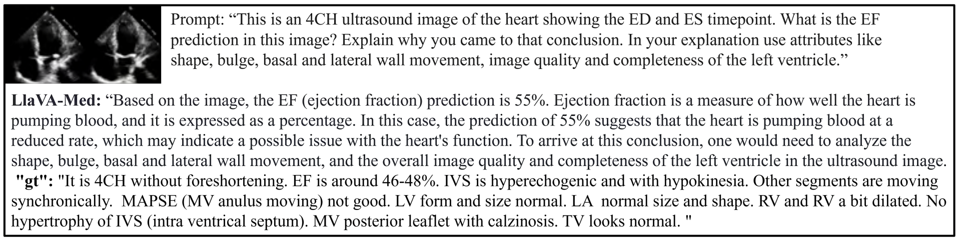
Model	bert	sbert	biobert	simcse	medsbert	rouge	rougeL	bleu	wcount
LlaVA-Med	0.81	0.49	0.53	0.69	0.95	0.25	0.15	0.01	541
NLE-EF-13B	0.88	0.56	0.63	0.75	0.96	0.26	0.15	0.00	194
NLE-EF-7B	0.88	0.56	0.63	0.75	0.95	0.25	0.15	0.00	207
NLE-EF-13B self-instr	0.84	0.54	0.59	0.75	0.95	0.21	0.12	0.00	541
NLE-EF-7B self-instr	0.85	0.57	0.62	0.75	0.95	0.19	0.10	0.00	622

Table 3. Additional metrics for the different NLE approaches. BERT, BioBERT, SimCSE are based on words, whereas sBERT,medsbert are on sentence level. BioBERT and medsbert are trained on clinical data. SimCSE is targeted for detecting contradictions. Rouge and BLEU are conventional NLP scores that are also dependent on the text length. Word count (wcount) accounts for the diversity of the different approaches.

3 Prompt examples

```
Example 1: Llama input prompt
{"instruction":
"Explain why the ejection fraction is estimated as 69%. ",
"input": "In the echocardiography image, it is measured that there is no septal bulge
present, the shape of the left ventricle looks normal, the movement of the segments
is normal, the image quality is reduced, the basal points move normal, the left
ventricle is cut, and the apex moves with a large amplitude. There might be
foreshortening. The bulge value is 417, the height over width value is 2.00, the
segment movement is normal, the apex moves 13.61%, the basal points move by 7.28%,
contrast is 4.76.",
"response": "The EF estimation in this case seems to be based
on several factors observed from ....}
```

```
Example 2: Prompt for the attribute Bulge for the Mistral model
<s> [INST] I have the following text that describes an image and I want you to answer
some questions about it by selecting one from different options. [/INST] Sure let me
help you with that, what is the text and the question </s> <s>[INST] This is the
text: There is a large bulge. Does the text mention any kind of bulge? Pick ONE
final answer out of: [1/bulge], [2/not specified in the text] or [3/none]. [/INST]
Final answer: [1/bulge] \n\n Explanation: The text states that there is a large bulge
present in the image, therefor it is present. </s><s> [INST] Great. Now do the
same task for the following text: TEXT_TO_INSERT Does the text mention any kind
of a bulge? Pick ONE final answer out of: [1/bulge], [2/not specified in the text]
or [3/none]. And please answer in the format:
Final answer: [option] \n\n Explanation: text [/INST]
```



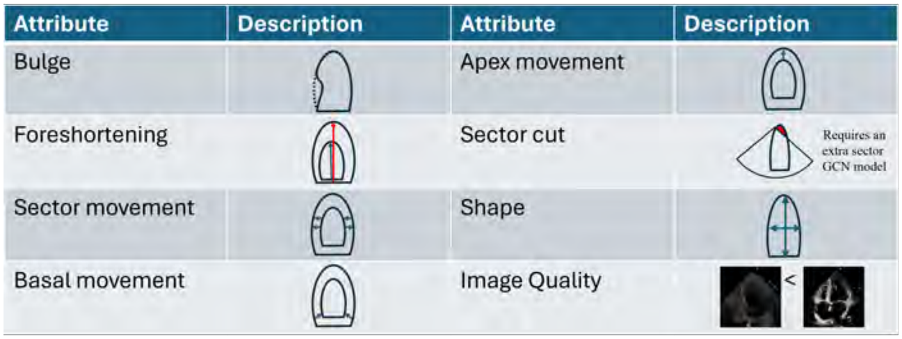
Prompt: "This is an 4CH ultrasound image of the heart showing the ED and ES timepoint. What is the EF prediction in this image? Explain why you came to that conclusion. In your explanation use attributes like shape, bulge, basal and lateral wall movement, image quality and completeness of the left ventricle."

LlaVA-Med: "Based on the image, the EF (ejection fraction) prediction is 55%. Ejection fraction is a measure of how well the heart is pumping blood, and it is expressed as a percentage. In this case, the prediction of 55% suggests that the heart is pumping blood at a reduced rate, which may indicate a possible issue with the heart's function. To arrive at this conclusion, one would need to analyze the shape, bulge, basal and lateral wall movement, and the overall image quality and completeness of the left ventricle in the ultrasound image.

"gt": "It is 4CH without foreshortening. EF is around 46-48%. IVS is hyperechogenic and with hypokinesia. Other segments are moving synchronically. MAPSE (MV anulus moving) not good. LV form and size normal. LA normal size and shape. RV and RV a bit dilated. No hypertrophy of IVS (intra ventrical septum). MV posterior leaflet with calzinosis. TV looks normal. "

Fig. 1. Prompt example used in LlaVA-Med. To compare with the NLE-EF output, we present here the same most-left example shown in Fig.2 in the main paper.

4 Attributes



Attribute	Description	Attribute	Description
Bulge		Apex movement	
Foreshortening		Sector cut	Requires an extra sector GCN model
Sector movement		Shape	
Basal movement		Image Quality	<

Fig. 2. Overview on attributes: Illustrative description of the different attributes used as input for the text model. Image Quality and sector cut are based on the image.