EchoMEN: Combating Data Imbalance in Ejection Fraction Regression via Multi-Expert Network

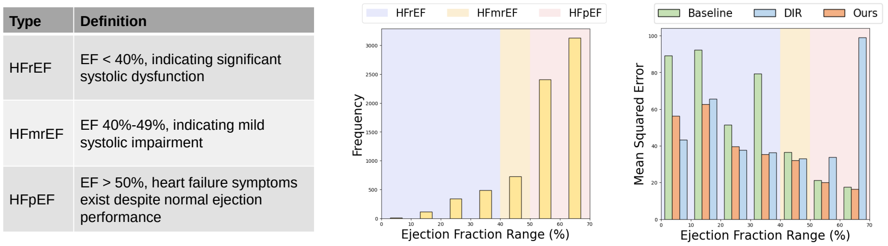
Song Lai[1,2], Mingyang Zhao[2], Zhe Zhao[1], Shi Chang[5], Xiaohua Yuan[5], Hongbin Liu[2,3], Qingfu Zhang[1], and Gaofeng Meng[2,3,4⋆]

[1] City University of Hong Kong, Hong Kong SAR
songlai2-c@my.cityu.edu.hk,qingfu.zhang@cityu.edu.hk
[2] Centre for Artificial Intelligence and Robotics, HK Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR
[3] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
gfmeng@nlpr.ia.ac.cn
[4] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[5] Central South University, Changsha, China

Abstract. *Ejection Fraction* (EF) regression faces a critical challenge due to severe *data imbalance* since samples in the normal EF range significantly outnumber those in the abnormal range. This imbalance results in a bias in existing EF regression methods towards the normal population, undermining *health equity*. Furthermore, current imbalanced regression methods struggle with the head-tail performance trade-off, leading to increased prediction errors for the normal population. In this paper, we turn to ensemble learning and introduce EchoMEN, a multi-expert model designed to improve EF regression with balanced performance. EchoMEN adopts a two-stage decoupled training strategy. The first stage proposes a Label-Distance Weighted Supervised Contrastive Loss to enhance representation learning. This loss considers the label relationship among negative sample pairs, which encourages samples further apart in label space to be further apart in feature space. The second stage trains multiple regression experts independently with variably re-weighted settings, focusing on different parts of the target region. Their predictions are then combined using a weighted method to learn an unbiased ensemble regressor. Extensive experiments on the EchoNet-Dynamic dataset demonstrate that EchoMEN outperforms state-of-the-art algorithms and achieves well-balanced performance throughout all heart failure categories. Code: https://github.com/laisong-22004009/EchoMEN.

Keywords: Echocardiography · Ejection Fraction · Data Imbalance · Multi-Expert Network.

⋆ Corresponding author



Type	Definition
HFrEF	EF < 40%, indicating significant systolic dysfunction
HFmrEF	EF 40%-49%, indicating mild systolic impairment
HFpEF	EF > 50%, heart failure symptoms exist despite normal ejection performance

(a) Definition of Heart Failure **(b) Ejection Fraction Distribution** **(c) Mean Squared Error Distribution**

Fig. 1. (a) defines different types of heart failure. (b) illustrates the distribution of EF labels within the EchoNet-Dynamic dataset, where HFpEF is the most prevalent category, followed by HFmrEF, and HFrEF being the least common. (c) demonstrates the distribution of Mean Squared Error (MSE) for baseline, DIR and EchoMEN.

1 Introduction

Ejection Fraction (EF) is a critical indicator of heart systolic function and is crucial for diagnosing heart failure [3,9]. Recent advances in deep learning techniques have significantly improved the performance of EF regression. For instance, EchoNet [13] utilized an R2+1D ResNet architecture for end-to-end EF prediction, while EchoCoTr [12] adopted a vision transformer for this task. EchoGNN [11] introduced an explainable EF framework through Graph Neural Networks. Despite these progress, a critical challenge persists – the impact of imbalanced data. According to [10], heart failure can be classified into three categories based on EF measurements: Reduced EF (HFrEF), Mildly Reduced EF (HFmrEF), and Preserved EF (HFpEF). The prevalence of data from normal individuals leads to a model bias towards the HFpEF category, consequently increasing the prediction error for HFrEF and HFmrEF cases, as summarized in Fig. 1. This imbalance poses a significant challenge to health equity, as accurate diagnosis across all categories of heart failure is crucial for effective treatment planning and health outcomes.

Yang *et.al* [20] first investigated the problem of learning from imbalanced data with continuous targets, which we refer to as Deep Imbalanced Regression (DIR). They proposed a solution that performs distribution smoothing on both label and feature spaces, demonstrating potential in handling imbalanced data scenarios. However, when directly applied to the EF estimation task, their method faces the following challenges: (1) Despite alleviating the impact of data imbalance on the model to some extent, it lacks sufficient consideration of the distribution of sample labels in EF estimation. This leads to limited representation capability in scenarios where the label semantic distribution is extremely imbalanced. (2) It adopts a single model structure for modeling and representation learning, lacking the ability to specifically model samples from different EF ranges. Consequently, it fails to effectively capture and represent the unique patterns exhibited by samples across different EF ranges. As shown in Fig. 1(c), the DIR method reduces errors in HFrEF but *increases the error of HFpEF*.

To address these challenges, we leverage contrastive learning and ensemble learning to consider continuous label relationships and mitigate the negative impact between head and tail data, respectively. In this paper, we propose a method termed EchoMEN to combat data imbalance in EF regression via a Multi-Expert Network. Firstly, to capture the distance relationship between data points, we enhance the encoder's feature learning through a Label-Distance Weighted Supervised Contrastive Loss ($\mathcal{L}_{\text{LDW-SupCon}}$), optimizing for dissimilarity among negative sample pairs by introducing an extra label-space continuity constraint. Secondly, to structurally alleviate the head-tail performance trade-off, we design multiple regression experts and an expert aggregator. The regression experts are independently trained, with each expert focusing on distinct aspects of the target distribution via a Re-weighted Regression Loss ($\mathcal{L}_{\text{R-Regression}}$). Subsequently, the aggregator combines all expert outputs to generate an unbiased final prediction. We validate our proposed method on the benchmark EchoNet-Dynamic dataset and demonstrate its superiority over baseline approaches.

Our contributions are: (1) To the best of our knowledge, we are the first work dedicated to addressing the imbalance issue in EF regression and enhancing the fairness of heart failure diagnosis. (2) We propose a supervised contrastive loss $\mathcal{L}_{\text{LDW-SupCon}}$ to boost representation learning in imbalanced dataset, which leverages the additional information encoded in label-space relationships. (3) We introduce the concept of ensemble learning to the domain of EF estimation and design a multi-expert model, which structurally alleviates the trade-off of head-tail performance (4) EchoMEN surpasses baseline approaches with higher estimation accuracy and delivers a more balanced performance.

2 Methodology

We denote $\mathcal{D} : \{(x_i, y_i)\}_{i=1}^{N}$ as the training dataset consisting of N instances. Each instance i comprises an echocardiogram video $x_i \in \mathbb{R}^{T \times W \times H \times C}$ and a ground truth $y_i \in \mathbb{R}$. As illustrated in Fig. 2, EchoMEN employs a two-stage decoupled training scheme, *Representation Learning* and *Multiple Experts Learning*, to address the challenge of data imbalance in EF regression. The details of each stage are described as follows.

2.1 Label-Distance Weighted Supervised Contrastive Loss

In this stage, we aim to learn a feature embedding network f_θ from labelled dataset $\mathcal{D}$. Motivated by the concept of distribution continuity [20], we integrate it with contrastive learning and introduce a novel contrastive loss $\mathcal{L}_{\text{LDW-SupCon}}$ to enhance representation learning in imbalanced data scenarios. Specifically, let $\mathcal{I}$ denote the sample indices for a randomly sampled batch during training, $\mathcal{P}(i) := \{j \in \mathcal{I} \mid y_j = y_i \wedge j \neq i\}$ as the set of indices for all positive samples

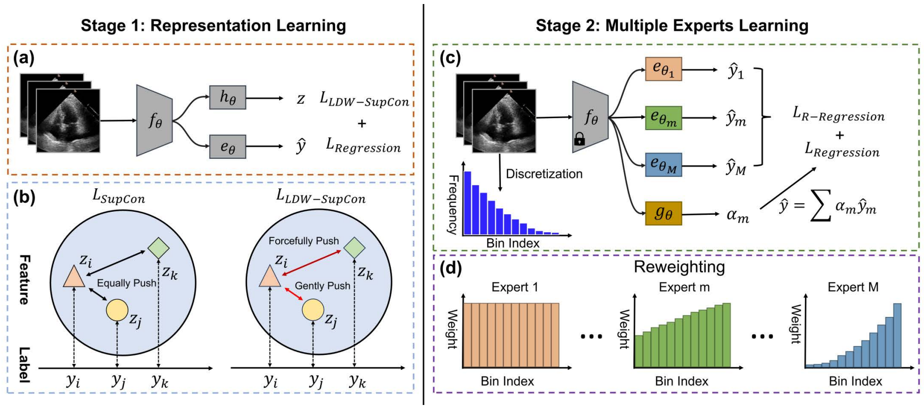


Fig. 2. Overview of EchoMEN. EchoMEN is trained using a two-stage decoupled training approach. Stage 1: (a) The video encoder is optimized with a *Label-Distance Weighted Supervised Contrastive Loss* ($\mathcal{L}_{\text{LDW-SupCon}}$) and a naive regression loss. (b) The $\mathcal{L}_{\text{LDW-SupCon}}$ enhances representation learning by differentially treating negative pairs based on label distances. Stage 2: (c) Multiple regression experts are trained using individual *Re-weighted Regression Loss* and are combined with an expert aggregator. (d) The varying sample weight distributions adopted by different experts reflect the diversity in balancing, moving from treating all samples equally to focusing more on sparsely distributed areas. The expert aggregator votes on the predictions from different experts and ultimately outputs the final outcome.

for anchor sample i and $\mathcal{Q}(i) := \{j \in \mathcal{I} \mid y_j \neq y_i\}$ as the negative samples, then $\mathcal{L}_{\text{LDW-SupCon}}$ can be formally expressed as:

$$\mathcal{L}_{\text{LDW-SupCon}} = -\sum_{i \in \mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p \in \mathcal{P}(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{j \in \mathcal{Q}(i)} \beta_{ij} \exp(\mathbf{z}_i \cdot \mathbf{z}_j / \tau)} \tag{1}$$

where $\mathbf{z}_i = h(f(x_i))$ denotes the normalized feature vector of sample i with h as the projector network. τ represents a temperature scaling parameter, and the dot product $\mathbf{z}_i \cdot \mathbf{z}_j$ measures the similarity between the embeddings of the sample i and j. β_{ij} is calculated as a normalized weight reflecting the distance d_{ij} between the labels of the anchor sample i and a negative sample j, defined by:

$$\beta_{ij} = \frac{|\mathcal{Q}(i)| \cdot \exp(d_{ij})}{\sum_{k \in \mathcal{Q}(i)} \exp(d_{ik})} \tag{2}$$

In practice, d_{ij} can be simply instantiated as L_1 or L_2 label difference.

Our proposed $\mathcal{L}_{\text{LDW-SupCon}}$ can be interpreted as a generalization of supervised contrastive loss [6] $\mathcal{L}_{\text{SupCon}}$ from pushing all negative samples equally to considering the inherent continuity underlying the labels in regression problems. As illustrated in Fig. 2(b), given randomly sampled indices i, j, k with label $y_i < y_j < y_k$, their corresponding feature embeddings are z_i, z_j, z_k respectively.

For anchor sample i, $\mathcal{L}_{\text{SupCon}}$ minimizes the similarity between negative pairs $\mathbf{z}_i \cdot \mathbf{z}_j$ and $\mathbf{z}_i \cdot \mathbf{z}_k$ and assigns them equal weights. However, our $\mathcal{L}_{\text{LDW-SupCon}}$ uses weight term β to differentially treat negative pairs, which promotes a larger feature space separation of $\mathbf{z}_i \cdot \mathbf{z}_k$ than $\mathbf{z}_i \cdot \mathbf{z}_j$. By introducing an extra label continuity contrastive constraint, $\mathcal{L}_{\text{LDW-SupCon}}$ can improve the representational capabilities of the video encoder. The continuity of the feature space can align more effectively with the target space, thereby aiding EchoMEN in combatting data imbalance in EF regression.

Moreover, to accelerate the learning of video encoder, we further adopt a regression loss $\mathcal{L}_{\text{Regression}} = \frac{1}{N}\sum_{i}^{N} L(e(f(x_i)), y_i)$, where e is an extra regression head. Thus, the overall loss function for *Representation Learning* stage is $\mathcal{L}_{\text{stage1}} = \mathcal{L}_{\text{Regression}} + \lambda\mathcal{L}_{\text{LDW-SupCon}}$, where λ is a coefficient that controls the relative importance of the two tasks.

2.2 Multiple Experts Training

Based on the recent studies [18,7,19] of using *Multi-Expert Networks* in long-tailed *classification* tasks, we extend it to the domain of imbalanced EF *regression*. The key to such methods lies in ensuring diversity among different experts. In long-tailed classification, diversity is achieved by re-sampling or dividing the dataset into different subsets. However, experimental results in Tab. 1 show that such approaches perform poorly in the EF regression problem, leading to a situation where, although each expert performs well within their focused region, aggregating the results of experts proves to be difficult. To address this issue, we introduce a re-weighting (cost-sensitive) learning strategy to enhance correlation among experts.

As presented in Fig. 2(c), our multiple-expert architecture comprises a shared video encoder f_θ, M independent regression experts $E = \{e_{\theta_1}, ..., e_{\theta_M}\}$, and an expert aggregator g_θ. This design has the advantage that it is computationally lightweight and makes all heads rely on a common feature embedding z. In this stage, we fix the parameter of f_θ and focus on the training of heads. We also introduce B equally spaced bins across the target range and compute the count of data points per bin, denoted by $\mathbf{f} = (f_1, \cdots, f_B)$.

To achieve diversity and enhance complementarity among experts, we train each expert e_{θ_m} with different *Re-weighted Regression Loss* $\mathcal{L}^{m}_{\text{R-Regression}}$, which is defined by:

$$\mathcal{L}^{m}_{\text{R-Regression}} = \sum_{i=1}^{N} w_i^m L(\hat{y}_i^m, y_i) \tag{3}$$

where $\hat{y}_i^m$ represents the prediction of sample i from the expert e_{θ_m}. w_i^m indicates the weight of sample i when calculating regression loss, expressed as follow:

$$w_i^m = \left(f_{b(i)}\right)^{-p_m}, \quad \text{with} \quad p_m = \frac{m}{M-1}, m \in \{0, ..., M-1\} \tag{4}$$

where $b(i)$ indicates the bin of sample i. The parameter p_m modulates the focus intensity of an expert on less populated regions. Fig. 2(d) displays the variation

Table 1. Quantitative comparisons on the EchoNet-Dynamic dataset. The best two results are indicated using bold fonts and underlining, separately.

Metrics	MAE ↓				GM ↓			
Shot	All	HFrEF	HFmrEF	HFpEF	All	HFrEF	HFmrEF	HFpEF
SQInv [16]	5.05	7.28	6.34	4.52	3.27	4.70	4.16	2.99
DIR [20]	4.68	5.99	5.33	4.51	2.96	3.68	3.75	2.90
Focal-R [8]	4.13	6.36	5.76	3.56	2.57	3.95	3.75	2.30
RankSim [2]	3.97	6.12	5.44	**3.43**	2.49	3.95	3.72	2.22
EchoNet [13]	4.05	6.26	5.96	3.45	2.51	3.74	3.91	2.23
Esfeh et.al [5]	4.46	6.67	6.52	3.97	2.98	4.72	4.92	2.59
Reynaudfor et.al [15]	5.42	10.13	9.91	4.01	3.40	6.50	7.94	2.74
EchoGNN [11]	4.45	6.83	5.93	3.93	2.85	4.67	3.83	2.54
EchoCoTr [12]	3.96	6.02	5.31	3.47	2.52	3.87	3.72	2.24
EchoMEN(Ours)	**3.93**	**5.94**	**5.29**	3.44	**2.44**	**3.64**	**3.71**	**2.17**

of sample weight distribution across bins with an increase of p_m. When $p_m = 0$, experts treat each sample equally; when $p_m = 1$, experts adopt inverse frequency weighting and fully compensate for density variations in the dataset. This setting of multi-expert training allows each expert to consider samples from other areas during learning and reduces the error caused by incorrect expert aggregation.

To combine the predictions of individual ensemble members, aggregator g_θ takes z as input and compute a set of weights $\alpha_m = h(f(x_i))$, each corresponding to one of the M experts. The final prediction of EchoMEN is $\hat{y}_i = \sum_{m=1}^{M} \alpha_m \cdot \hat{y}_i^m$. Here, our aggregator is trained in an end-to-end fashion, using the regression loss $\mathcal{L}_{\text{Regression}} = \frac{1}{N} \sum_{i=1}^{N} L(\hat{y}_i, y_i)$.

3 Experiments

We conduct extensive experiments to validate the proposed method for EF regression and compare it with competing approaches using benchmark datasets.

3.1 Dataset and Evaluation Metrics

The EchoNet-Dynamic dataset [13] comprises 10,030 videos, each annotated with a EF value. Each video is a sequence of 112×112 grayscale images, capturing the dynamic motion of the heart across different cardiac cycles. The dataset is divided into training, validation, and testing splits containing 7,460, 1,288, and 1,277 videos, respectively. The evaluation metrics include both *Mean Absolute Error* (MAE) and *Geometric Mean* (GM). Experiments are performed on both the overall dataset and its three subsets.

Table 2. Ablation Study on EchoNet-Dynamic

Metrics	MAE ↓				GM ↓			
Shot	All	HFrEF	HFmrEF	HFpEF	All	HFrEF	HFmrEF	HFpEF
w/o $\mathcal{L}_{\text{LDW-SupCon}}$	4.03	6.33	5.33	$\underline{3.49}$	2.56	4.27	3.74	2.25
w/o multiple experts	4.02	6.50	5.31	$\underline{3.49}$	2.53	4.35	3.72	$\underline{2.23}$
w/o re-weighting	4.36	5.97	5.75	3.97	2.76	3.69	3.81	2.55
w/o aggregator	$\underline{4.01}$	$\underline{5.96}$	**4.99**	3.59	$\underline{2.49}$	$\underline{3.66}$	**3.40**	2.26
Ours	**3.93**	**5.94**	$\underline{5.29}$	**3.44**	**2.44**	**3.64**	$\underline{3.71}$	**2.17**

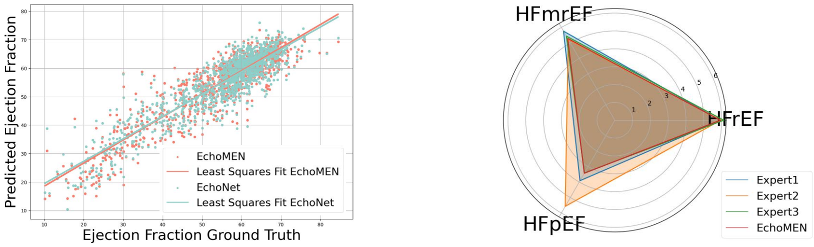
3.2 Implementation Details

We implement the proposed EchoMEN model using the PyTorch library [14]. The video encoder f_θ utilizes a R(2+1)D ResNet [17] pretrained on the Kinetics-400 dataset [4]. Each expert e_θ is a single linear layer and g_θ is made of linear layers followed by Softmax for outputting weights. h_θ is a two-layer MLP with ReLU. For label discretization, the length of f_b is set to 1. All experiments are conducted on a NVIDIA GeForce GTX A100 GPU, with a batch size of 20. Our EchoMEN model is optimized using Stochastic Gradient Descent (SGD) [1] with a learning rate of 10^{-4} and a decay factor of 0.1 every 10 epochs. Stage 1 is trained for 40 epochs with λ of 0.5 and stage 2 has 20 epochs. MSE is employed as the regression loss function.

3.3 Comparisons

We compare EchoMEN with the following recent EF regression baselines: (a) EchoNet [13]; (b) Esfeh *et.al* [5]; (c) Reynaudfor *et.al* [15]; (d) EchoGNN [11]; (e) EchoCoTr [12]. Additionally, we implement current deep imbalanced methods and apply them to EF regression for further comparison: (f) **SQInv** [16] utilizes a square-root-inverse-frequency weighting scheme; (g) **DIR** [20] performs distribution smoothing on label and feature spaces; (h) **Focal-R** [8] is the regression version of Focal Loss; (i) **RankSim** [2] adds a regularization term to leverage the continuity of targets.

The quantitative comparison results are reported in Tab. 1. As observed, EchoMEN outperforms all existing approaches in terms of overall performance, as indicated by superior results across both metrics. Notably, in sparsely distributed regions, including HFrEF and HFmrEF, our work achieves the best results compared to all competitors, highlighting that EchoMEN effectively utilizes the predictions of different experts to address the challenges of imbalanced datasets with significant frequency variations. As for the densely populated HFpEF subset, EchoMEN is on par with the existing state-of-the-art method [2] but with higher overall accuracy. In summary, our proposed method not only surpasses all competing methods on the entire dataset but also demonstrates a well-balanced performance across all three subsets.



(a) Comparison between EchoMEN and EchoNet **(b) Each expert and EchoMEN MAE**

Fig. 3. (a) compares the performance of EchoMEN and EchoNet. (b) showcases per-expert and aggregated MAE on EchoNet-Dynamic. EchoMEN nearly matches the performance of the best expert on each category.

Moreover, we present the predicted EF values of EchoMEN and EchoNet against the ground truth values in Fig. 3(a), where the lines indicate the least-squares regression line between the model predictions and the ground truth. It can be observed that the slope of EchoMEN is closer to 1 compared with EchoNet, suggesting a better fitting result.

Additionally, in Fig. 3(b), we provide an analysis of each expert used in our method. We observe that the aggregator plays a crucial role in dynamically weighting the outputs of different experts. The prediction quality of the ensemble method approaches that of an oracle, which consistently optimally weights the outputs of all experts across all subsets.

3.4 Ablation Study

We explore the influence of different design choices in our framework by training the following variants on the same training data. The variations include: (a) **w/o $\mathcal{L}_{\textbf{LDW-SupCon}}$** utilizes a standard supervised contrastive loss; (b) **w/o multiple experts** removes the multiple experts learning stage; (c) **w/o re-weighting** divides the dataset into multiple subsets and trains regression experts using different subset; (d) **w/o aggregator** replaces the aggregator-based weighted mechanism with average voting.

Tab. 2 showcases the impact of removing one of the components of EchoMEN in the EchoNet-Dynamic dataset. Overall, each ablation results in increases in MAE and GM and combining all designs leads to the best performance. Individually, $\mathcal{L}_{\text{LDW-SupCon}}$ and multiple experts primarily benefit the HFrEF cases, whereas re-weighting mainly decreases the error in the HFmrEF category. The aggregator makes the prediction error more balanced.

4 Conclusion and Future Work

This paper presented a multi-expert model termed EchoMEN to tackle the data imbalance in EF estimation. EchoMEN implements a decoupled learning strategy in two stages. The first stage introduces a contrastive loss $\mathcal{L}_{\text{LDW-SupCon}}$, which adaptively embeds target relationships into the objective to differently weight negative pairs, achieving better alignment between feature and label space for representation learning. The second stage trains multiple regression experts and weights all predictions for final estimation, allowing the strengths of each expert to be leveraged within their respective target areas. Experiments on the EchoNet-Dynamic dataset demonstrate the effectiveness of EchoMEN.

Limitations and Future Work: EchoMEN requires computing outputs of all experts during inference, which leads to significant computational cost with an increasing number of experts. Therefore, we plan to refine our multi-expert architecture with dynamic routing to decrease complexity by selectively activating experts only when needed.

Acknowledgments. This research was partly supported by the National Natural Science Foundations of China (Grants No.62376267), the Key Basic Research Foundation of Shenzhen, China (JCYJ20220818100005011) and the innoHK project.

Disclosure of Interests. The authors declare no competing interests.

References

1. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers. pp. 177–186. Springer (2010) 7
2. Gong, Y., Mori, G., Tung, F.: Ranksim: Ranking similarity regularization for deep imbalanced regression. In: International Conference on Machine Learning. pp. 7634–7649. PMLR (2022) 6, 7
3. Huang, H., Nijjar, P.S., Misialek, J.R., Blaes, A., Derrico, N.P., Kazmirczak, F., Klem, I., Farzaneh-Far, A., Shenoy, C.: Accuracy of left ventricular ejection fraction by contemporary multiple gated acquisition scanning in patients with cancer: comparison with cardiovascular magnetic resonance. Journal of Cardiovascular Magnetic Resonance **19**(1), 1–9 (2017) 2
4. Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., et al.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017) 7
5. Kazemi Esfeh, M.M., Luong, C., Behnami, D., Tsang, T., Abolmaesumi, P.: A deep bayesian video analysis framework: towards a more robust estimation of ejection fraction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 582–590. Springer (2020) 6, 7
6. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020) 4
7. Li, Y., Wang, T., Kang, B., Tang, S., Wang, C., Li, J., Feng, J.: Overcoming classifier imbalance for long-tail object detection with balanced group softmax. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10991–11000 (2020) 5

8. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: International Conference on Computer Vision (2017) 6, 7
9. Loehr, L.R., Rosamond, W.D., Chang, P.P., Folsom, A.R., Chambless, L.E.: Heart failure incidence and survival (from the atherosclerosis risk in communities study). The American journal of cardiology **101**(7), 1016–1022 (2008) 2
10. Members:, A.F., McDonagh, T.A., Metra, M., Adamo, M., Gardner, R.S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., et al.: 2021 esc guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the european society of cardiology (esc). with the special contribution of the heart failure association (hfa) of the esc. European journal of heart failure **24**(1), 4–131 (2022) 2
11. Mokhtari, M., Tsang, T., Abolmaesumi, P., Liao, R.: Echognn: Explainable ejection fraction estimation with graph neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 360–369. Springer (2022) 2, 6, 7
12. Muhtaseb, R., Yaqub, M.: Echocotr: Estimation of the left ventricular ejection fraction from spatiotemporal echocardiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 370–379. Springer (2022) 2, 6, 7
13. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020) 2, 6, 7
14. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019) 7
15. Reynaud, H., Vlontzos, A., Hou, B., Beqiri, A., Leeson, P., Kainz, B.: Ultrasound video transformers for cardiac ejection fraction estimation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2021), `https://api.semanticscholar.org/CorpusID:235727599` 6, 7
16. Steininger, M., Kobs, K., Davidson, P., Krause, A., Hotho, A.: Density-based weighting for imbalanced regression. Machine Learning **110**, 2187–2211 (2021) 6, 7
17. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6450–6459 (2018) 7
18. Wang, X., Lian, L., Miao, Z., Liu, Z., Yu, S.: Long-tailed recognition by routing diverse distribution-aware experts. In: International Conference on Learning Representations (2020) 5
19. Xiang, L., Ding, G., Han, J.: Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. pp. 247–263. Springer (2020) 5
20. Yang, Y., Zha, K., Chen, Y., Wang, H., Katabi, D.: Delving into deep imbalanced regression. In: International Conference on Machine Learning. pp. 11842–11851. PMLR (2021) 2, 3, 6, 7