Explainable Integrative Bipartite Graph Convolutional Neural Network for Predicting Ejection Fraction in Echocardiography

Seungeun Lee[1], Jaeyoung Kim[2], Kyungtae Kang[1(✉)], and Mingon Kang[3(✉)]

[1] Department of Computer and Information Security, Hanyang University, Ansan, South Korea
[2] Department of Convergence Medicine, Korea University Hospital, Ansan, South Korea
[3] Department of Computer Science, University of Nevada Las Vegas, Nevada, USA
ktkang@hanyang.ac.kr, mingon.kang@unlv.edu

Abstract. Ejection fraction (EF) estimation in echocardiography is a key indicator to examine cardiac functions and to determine the optimal treatments for patients prone to heart dysfunctions, such as heart failure. Recently, machine learning has shown promising predictive performance, as diagnostic tools, to estimate EF using echocardiograms. However, most state-of-the-art models have overlooked diversity of phenotypes in echocardiography, derived from patient's demography (e.g., sex and age). In this study, we propose a novel integrative bipartite graph neural network (IBi-GNN) that integrates demographic variables of patients with echocardiograms to improve the EF predictive performance and model interpretability in precision medicine. In the experiments, IBi-GNN significantly reduced the estimation errors compared to the benchmark models, and the significant improvement was statistically assessed. We also show that IBi-GNN is interpretable to identify interaction between the multi-modalities. The interpretation provides comprehensive understanding of the relationships between demographic factors and cardiac structures. The open-source codes are publicly available at https://github.com/datax-lab/IBi-GNN.

Keywords: Explainable AI · Demographic interactions · Integrative Bipartite Graph CNN · Echocardiography

1 Introduction

Accurate estimation of left ventricular ejection fraction (EF) in echocardiography is essential to make the optimal treatment recommendations for diagnosis, prognosis, and treatment of patients with heart failure [18]. The EF measurements categorize heart failure types into three: (1) heart failure with reduced EF (HFrEF, EF $<$ 40), (2) heart failure with mid-range reduced EF (HFmEF, 40$\leq$EF$<$50), and (3) heart failure with preserved EF (HFpEF, EF$\geq$50). In particular, patients with heart failure with reduced ejection fraction (HFrEF) con-

stitute approximately 50% of all heart failure cases, and they are more likely to have coronary heart disease or chronic kidney disease [11].

Echocardiography is the most common modality to measure ejection fraction, due to the advantages of its low cost, utility, and safety [15]. An echocardiogram creates temporal pictures of the beating hearts using sound waves in echocardiography. Ejection fractions can be estimated by characterizing left ventricular volume difference at end-systole (ES) and end-diastole (ED). However, manual estimation of EF using echocardiograms is a tremendously time-consuming task involving frame detection and left ventricular volume tracing, which is prone to human error.

Recently, deep learning has succeeded in automatic analysis of echocardiograms, for deep learning is capable of capturing spatio-temporal patterns from videos. A number of deep learning models has improved the predictive performance of ejection fraction estimation. A multi-task deep learning model analyzed echocardiogram using left ventricle segmentation and landmark detection, reflecting the conventional clinical workflow for ejection fraction calculation [10]. A 3D Convolutional Neural Network (3D-CNN) captures the spatio-temporal patterns in echocardiogram video clips [4, 6, 14]. A transformer model efficiently incorporated temporal patterns over the entire frames [13]. A graph neural network (GNN) learned the relationships between echocardiogram frames to estimate EF from echocardiogramic videos, identifying important frames for the estimation [12].

Understanding demography-based spatiotemporal patterns in echocardiography is critical to improve ejection fraction estimation in precision medicine. One of the conventional approaches is data integration of demographic variables on echocardiograms. A number of deep learning studies have integrated tabular data (e.g., demographic/clinical variables) with medical imaging data. A transformer fusion model developed a unified integration approach by selectively focusing on relevant variables in the integration of each multi-modal data [25]. Contrastive learning-based models aligned multi-modality within the same samples, by maximizing their similarity in the latent feature space [7, 21]. A GNN-based model integrated multi-modal data using a graph for survival prediction [5]. However, such integrative approaches focus on combining the multi-modal information rather than identifying their interaction effects. Furthermore, to the best of our knowledge, there is no computational tool that predicts ejection fractions based on the interaction effects with demographic variables in echocardiography, although a number of literature has addressed the patients' demographic effects on cardiac structures and functions on echocardiography [3, 17].

In this study, we propose a novel integrative deep learning model (named IBi-GNN) that characterizes interaction effects between patients' demographic factors and echocardiogramic visual features and improves the predictive performance. IBi-GNN identifies demography-specific echocardiogramic features which are non-linearly associated to clinical outcomes (e.g., ejection fraction). The main contributions of the study are: (1) we present a novel integrative deep learning model, IBi-GNN, designed for identifying demography-specific visual features,

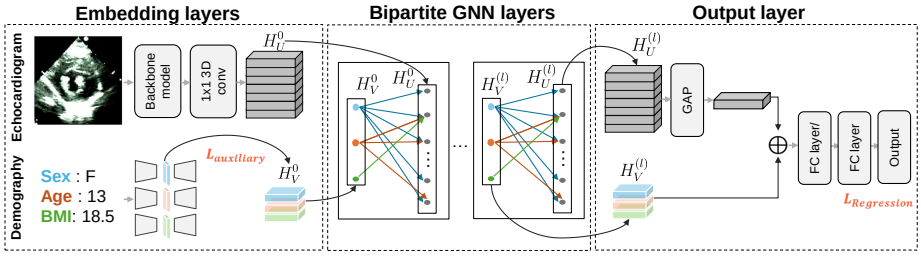

Fig. 1: IBi-GNN consists of: (1) the video embedding layers from echocardiograms, (2) the demography embedding layers, (3) the bipartite graph layers, and (4) the output layers. The embedding layers extract latent variables of each modality, and the bipartite graph layers captures their interactions.

via bipartite graph layers, (2) IBi-GNN showed statistically significant improvement on the EF estimation using echocardiograms in the experiments, and, (3) the identified demography-specific patterns can be leveraged, as new clinical knowledge, to make personalized treatments in precision medicine.

The next sections are organized as follows. The details of the proposed method are described in Section 2. The dataset and the experimental results are presented in Section 3. The interpretation and analysis of this study are discussed in Section 4. Finally, conclusions are presented.

2 Method

We propose a bipartite graph neural network that integrates echocardiogram and demographic variables to estimate ejection fraction by characterizing their interactions. Let an echocardiogram be $\mathfrak{E} \in \Re^{T \times H \times W}$ for the T frames of $W \times H$-sized color images, and demographic factors $\mathfrak{D} \in \Re^{P_{\mathfrak{D}}}$, where $P_{\mathfrak{D}}$ is the number of demographic variables. In this study, the demographic variables include age, sex, and Body Mass Index (BMI), calculated with height and weight. The embedding layers extract latent variables of each modality, and the bipartite graph layers captures their interactions. The bipartitie graph layers in the model can be trained to reflect interaction between demographic variable and echocardiogram features.

2.1 Model design

The architecture of the proposed model consists of the following layers: (1) embedding layers for an echocardiogram, (2) embedding layers for demographic factors, (3) bipartite graph layers, and (4) output layers (Fig. 1). The embedding layers extract video/demographic latent features from the two data modalities separately. The embedding layers for an echocardiogram generate the video latent feature. A video backbone model (e.g., video vision transformer [1], 3D-ResNet [9]) extracts video features. Then, time invariant spatial features are

generated by 1×1 3D convolution with d dimensions. The time invariant spatial features are represented by $\mathbf{H}_U = \{u_i | 1 \leq i \leq Q_{\mathfrak{E}}\}$, where $u_i \in \Re^{1 \times d}$ is i-th features of an echocardiogram, and $Q_{\mathfrak{E}}$ is the number of the low-ranked time invariant spatial features.

The embedding layers for demographic variables encode each variable to a latent vector though data reconstruction modules. The modules consist of encoder-decoder layers for the reconstruction of the input variables. The j-th demographic variable, denoted as $\mathfrak{D}_j$, is reconstructed to $\hat{\mathfrak{D}}$, by $v_j = Encoder(\mathfrak{D}_j)$ and $\hat{\mathfrak{D}}_j = Decoder(v_j)$. The encoder and decoder layers are comprised of fully connected layers with activation functions (i.e., autoencoder). Through the reconstruction learning, each demographic variable is encoded into $\mathbf{H}_V = \{v_j | j \in \{\text{age, bmi, sex}\}\}$, where $v_j \in \Re^{1 \times d}$ is the j-th demographic latent variable.

In the bipartite graph layers, the bipartite graph, $G = (\mathbf{U}, \mathbf{V}, \mathbf{E})$, consists of two distinct sets of nodes, $\mathbf{U}$ for video features and $\mathbf{V}$ for demographic features. $\mathbf{E} \subset U \times V$ is a set of edges represented by weighted adjacency matrix, $\mathbf{A} \in \Re^{(|U|+|V|)\times(|U|+|V|)}$. The proposed model considers one-way directed edges, only from the demographic set to the echocardiogramic set, which represent demographic interaction into echocardiograms as $A_{V \to U}$. The bipartite graph layers consist of L layers, each layer propagates the embeddings through graph convolutional networks (GCN) [8]. Let $H_u^{(l)}$ and $H_v^{(l)}$ $(0 \leq l \leq L)$ be the l-th hidden embedding for the nodes of U and V, respectively. $H_u^{(l)}$ and $H_v^{(l)}$ are represented as:

$$\begin{bmatrix} H_V^{(l+1)} \\ H_U^{(l+1)} \end{bmatrix} = \sigma \left(\begin{bmatrix} 0_{(V,V)} & \hat{A}_{(V \to U)} \\ 0_{(U \to V)} & 0_{(U,U)} \end{bmatrix} \begin{bmatrix} H_V^{(l)} \\ H_U^{(l)} \end{bmatrix} W^{(l)} + \begin{bmatrix} H_V^{(l)} \\ H_U^{(l)} \end{bmatrix} \Theta^{(l)} \right), \quad (1)$$

where $H_U^{(0)}$ and $H_V^{(0)}$ are the video and demographic embedding features respectively, which are introduced from the embedding layers. $\hat{A}$ is the normalized adjacency matrix, W^l and Θ^l learnable weight parameters, and σ is an activation function (i.e., ReLU in this study).

The output layer estimates ejection fractions from the video/demographic hidden embedding. The video hidden embedding, $H_U^{(L)}$, produces a vector, h_u, by the global average pooling layer. The video feature, h_u, is stacked to the demographic feature matrix, $H_v^{(L)}$, resulting in the output matrix, $\mathbf{G}$. Then, an ejection fraction is computed by through fully-connected layers.

2.2 Optimization

IBi-GNN optimizes the parameters by simultaneously minimizing both the regression loss and the reconstruction loss, as defined by the total loss function: $L^{\text{total}} = L^{\text{regression}} + \lambda L^{\text{auxiliary}}$. The regression loss uses weighted mean squared error (MSE) by kernel density estimation in (2), due to the data imbalance in terms of ejection fraction rates [20]. Each sample is weighted by the kernel density function ($f_w(\cdot)$) and hyper-parameter (α) that determines the degree of

weighting. In the mean time, the auxiliary reconstruction loss optimizes the reconstruction of demographic embedding. The reconstruction loss takes the sum of the mean squared errors of the continuous variables and cross-entropy losses of the categorical variables in (3).

$$L^{\text{regression}} = \frac{1}{N}\sum_{i=1}^{N} f_w(\alpha, y_i)\,(y_i - \hat{y}_i)^2\,, \tag{2}$$

$$L^{\text{auxiliary}} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j\in\{\text{age,BMI}\}} (\mathfrak{D}_{j^i} - \hat{\mathfrak{D}_{j^i}})^2 - \sum_{k\in\{\text{sex}\}} \mathfrak{D}_{k^i}\log(p_k^i)\right). \tag{3}$$

where N is the number of data sample, y_i is a ground truth value, $\hat{y_i}$ is model estimation, and p^k is the predicted probability value.

3 Experimental Results

3.1 Dataset and data preprocessing

We used the publicly available EchoNet-Pediatric EF dataset [16], which consists of echocardiogram video clips from two standard views, parasternal short-axis (PSAX) and apical four chamber (A4C), along with demographic variables. In this study, we utilized only the PSAX view. The image resolution of the echocardiograms was 112 × 112 pixels, and the corresponding ejection fraction (EF) values were provided as labels. The demographic variables included sex (57% males, 43% females), age (10.16 ± 5.4 years), weight (42.43 ± 25.8 kg), and height (137.37 ± 41.6 cm). After filtering out with missing values, we used 4,294 samples, containing both video and demographic data. For the echocardiogram data, we performed a video sampling stratgy to ensure a fixed number of video frames from each echocardiogram, each consisting of one or more cardiac cycles, with the number of frames varying per cycle [13]. Specifically, We selected one frame every 4 frames from each video, resulting in a total of 36 frames per video. To preprocess the demographic variables, we considered BMI instead of weight and height separately, along with sex and age. Additionally, we removed outliers from the BMI, weight, and height variables using the interquartile range (IQR) method to enhance the data quality.

3.2 Experimental design and implementation

We compared the performance of our model with current state-of-the-art benchmark methods, repeating the experiments twenty times for reproducibility. In each experiment, the data is split to 80% for training, 10% for validation, and 10% for testing. We normalized data using z-score normalization with the train data, and then the parameters were applied to the validation and the test data. We considered contrastive integration (Contrastive) [7], transformer integration (Transformer) [25], and graph attention network integration (GAT) [5], as the

Table 1: The performance comparison with the benchmark models. The best results are in boldface. * and ** indicate statistical significance compared to the second best integrative models (Wilcoxon signed-rank test, p<0.05*, p<0.01**).

Modality	Method	MSE	MAE	R2
Echocardiogrm & Demography	IBi-GNN (Ours)	**29.31 ± 3.87****	**3.95 ± 0.18****	**0.73 ± 0.05***
	GAT [5]	31.50 ± 3.85	4.07 ± 0.18	0.67 ± 0.06
	Contrastive [7]	40.80 ± 4.70	4.50 ± 0.25	0.63 ± 0.10
	Transformer [25]	37.55 ± 5.07	4.44 ± 0.24	0.58 ± 0.09
Echocardiogram	EchoCoTr [13]	34.27 ± 4.12	4.44 ± 0.36	0.67 ± 0.04
	R(2+1)D [16]	30.39 ± 3.60	4.06 ± 0.14	0.70 ± 0.04

benchmark of the integrative approach. We also included recent models that predict EF using echocardiogram data only, such as the vision transformer-based model (EchoCoTr) [13] and a CNN-based model (R(2+1)D) [16], as baseline benchmark. For the implementation of Contrastive, we replaced the original 2D image backbone model with a video backbone model and further optimized the number of fully connected layers (i.e., 2) and the hidden dimensions (i.e., 128) to apply to the EF prediction using echocardiogram. For Transformer, two multimodal latent features were integrated through the cross-attention mechanism proposed in [25]. For GAT, two GAT layers with hidden dimension of 64 were considered in the graph neural network, enabling the capture of complex relational structures in the data [5]. For the fair comparison in the integrative benchmark models, we used EchoCoTr as the video backbone model, which demonstrated better performance compared to R(2+1)D. We optimized the model parameters and hyperparameters, including the number of layers, hidden dimensions, learning rate, optimizer, and batch size, to minimize loss on the validation data for all the benchmark models.

For IBi-GNN, we designed the architecture of the demographic embedding layers with fully connected layers with a hidden dimension of 128, and the bipartite GNN layers with 2-layers with hidden dimensions of 64 and 128, respectively. For ejection fraction estimation, we used two fully connected layers: the first with a hidden dimension of 512, followed by an output layer with a dimension of 1.

We trained the model using the Adam optimizer with emperically optimized hyper-parameters: a learning rate of 1e-4, first and second moment estimates of 0.9 and 0.999 respectively.

3.3 Results

We computed Mean Squared Error (MSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) to evaluate the ejection fraction prediction in the twenty experiments. The proposed model, IBi-GNN, outperformed the other benchmark models, showing the lowest MSE of 29.31±3.87, MAE of 3.95±0.18, and the highest R^2 of 0.73±0.05 (Table 1). IBi-GNN showed enhanced predictive performance compared to the benchmark models (GAT, Contrastive, and

Transformer) of the integrative approach, and its improvement was statistically assessed by Wilcoxon rank sum test ($p < 0.05$). IBi-GNN also showed the improved performance compared to the benchmark that used only echocardiogram.

4 Model interpretation

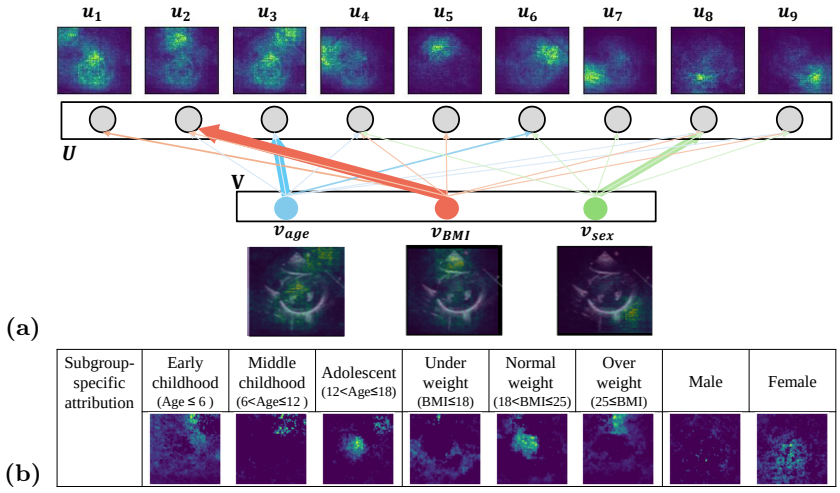

Fig. 2: (a) The edges of the bipartite GNN layers are visualized, reflecting demographic interactions with echocardiogram. The corresponding echocardiogram pattern of each video node vector u_i is highlighted. (b) Visualization of subgroup-specific attribution difference.

Ibi-GNN is an interpretable model that can capture relationships between demographic variable and echocardiogram features using bipartite GNN layers. For the model interpretation, we examined the bipartite graph layers learned with the training dataset. We computed importance scores of the edges of the bipartite graph layers by using GNN Explainer [24]. The interactions between demographic variables ($v_{age}, v_{BMI}, v_{sex}$) and the echocardiogramic visual features ($u_1, ..., u_9$) are illustrated in Fig. 2a, where the importance scores are represented by line weights. Visual feature maps on the nine echocardiogramic visual features (top) and three demographic variables (bottom) are visualized using DeepLift in Fig. 2a [19]. For instance, u_1 mainly highlights the top-left corner and centers, whereas u_2 are mainly associated to the top-middle and centers.

The learned bipartite graph shows that the echocardiogramic features are associated with the demographic variables with different weights. The identified visual features can be interpreted as demographic-specific features in the echocardiogram. Age, BMI, and sex each have distinct regions associated with ejection fraction. We discovered that u_2, u_3, and u_8 are mainly associated to

BMI, age, and sex, respectively, from the learned bipartite graph. For instance, the visual feature u_3, which are mainly associated with age, are highlighted in the center and top-right. The visual feature associated with BMI is highlighted in the center and top-middle corner. The visual feature associated with sex is highlighted in the bottom-middle region. The highlighted spots are well-aligned with clinical domain knowledge. In particular, the visual feature in the top-center corresponds to the area of epicardial adipose tissue at the echocardiogram, which is a significant adipose depot situated near the heart. This association between the epicardial adipose tissue and BMI has been reported [2]. Furthermore, our model identified the relationship of sex and the variations in both left ventricular and right ventricular volumes which underscore the nuanced influence of gender on cardiac structures.

To examine whether these regions differ across subgroups, we further visualized the subgroup-specific attribution differences by comparing the attribution value for each subgroup with the overall mean attribution value (Fig. 2b). For the analysis, we categorized the pediatric age groups into early childhood, middle childhood, and adolescence [23], and the BMI groups into underweight, normal weight, and overweight based on [22].

In the group of 12 years and older, echocardiographic features are highly scored in the centers on the heart's central regions—likely the left ventricle or interventricular septum—potentially reflecting increased myocardial mass or wall thickness associated with matured cardiac workload. In contrast, the 6–12 years old group and under 6 years old show a shift in emphasis toward the outer regions, such as the atria or peripheral ventricular zones, suggesting a transitional phase in cardiac development. Meanwhile, the group under 6 years old exhibits a relatively even distribution across the echocardiogram, displaying a notably uniform pattern indicative of an immature myocardium with minimal regional specialization.

For BMI-related regions, underweight individuals emphasize the lower outer myocardial regions, likely reflecting altered myocardial strain, due to reduced cardiac preload and ventricular wall stress. Those with a normal weight demonstrate a balanced distribution centered in the mid-myocardial regions, indicative of a normative myocardial workload and structural distribution. For individuals with a overweight, the concentration shifts to the upper outer myocardial regions, potentially corresponding to increased myocardial workload and altered hemodynamic stress patterns associated with excess adiposity, consistent with previously reported associations between obesity and regional myocardial dysfunction.

For sex-related regions, the female group exhibits more variation in the highlighted regions compared to the male group. This heterogeneity contrasts with the male group's consistent regional emphasis, suggesting uniform myocardial adaptation. In contrast, the greater variability observed in females may reflect differences in chamber sizes, wall thickness, or hemodynamic responses. Overall, IBi-GNN's model interpretation provided a comprehensive understanding of the intricate interplay between physiological factors and cardiac structures.

5 Conclusion

In this study, we propose a novel explainable integrative model, IBi-GNN, for ejection fraction estimation. IBi-GNN integrates echocardiograms and demographic features and learns their interactions using the bipartite graph neural network layers, which improve the predictive performance as well as the model interpretability. The optimized edges of the bipartite graph neural network layers can infer demography-specific echocardiogramic features for personalized diagnosis. Our study sheds light on compelling evidence linking demography features to echocardiogram patterns contributing to a comprehensive understanding of the intricate interplay between physiological factors and cardiac structures. The proposed model could be a useful personalized diagnosis tool for cardiovascular disease and will help one to understand the biological basis of cardiac phenotype.

Acknowledgments. This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP) - Innovative Human Resource Development for Local Intellectualization program (IITP-2025-RS-2020-II201741) and Artificial Intelligence Convergence Innovation Human Resources Development (Hanyang University ERICA)(No.RS-2022-00155885).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., Schmid, C.: Vivit: A video vision transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6836–6846 (2021)
2. Csecs, I., Czimbalmos, C., Toth, A., Dohy, Z., Suhai, I.F., Szabo, L., Kovacs, A., Lakatos, B., Sydo, N., Kheirkhahan, M., et al.: The impact of sex, age and training on biventricular cardiac adaptation in healthy adult and adolescent athletes: cardiac magnetic resonance imaging study. European Journal of Preventive Cardiology **27**(5), 540–549 (2020)
3. D'Andrea, A., Riegler, L., Rucco, M.A., Cocchia, R., Scarafile, R., Salerno, G., Martone, F., Vriz, O., Caso, P., Calabrò, R., et al.: Left atrial volume index in healthy subjects: clinical and echocardiographic correlates. Echocardiography **30**(9), 1001–1007 (2013)
4. Feng, Z., Sivak, J.A., Krishnamurthy, A.K.: Two-stream attention spatio-temporal network for classification of echocardiography videos. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1461–1465. IEEE (2021)
5. Fu, X., Patrick, E., Yang, J.Y., Feng, D.D., Kim, J.: Deep multimodal graph-based network for survival prediction from highly multiplexed images and patient variables. Computers in Biology and Medicine **154**, 106576 (2023)
6. Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J.H., Harrington, R.A., Liang, D.H., Ashley, E.A., Zou, J.Y.: Deep learning interpretation of echocardiograms. NPJ digital medicine **3**(1), 10 (2020)
7. Hager, P., Menten, M.J., Rueckert, D.: Best of both worlds: Multimodal contrastive learning with tabular and imaging data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23924–23935 (2023)

8. He, C., Xie, T., Rong, Y., Huang, W., Huang, J., Ren, X., Shahabi, C.: Cascade-bgnn: Toward efficient self-supervised representation learning on large-scale bipartite graphs. arXiv preprint arXiv:1906.11994 (2019)
9. Kataoka, H., Wakamiya, T., Hara, K., Satoh, Y.: Would mega-scale datasets further enhance spatiotemporal 3d cnns? arXiv preprint arXiv:2004.04968 (2020)
10. Li, H., Wang, Y., Qu, M., Cao, P., Feng, C., Yang, J.: Echoefnet: Multi-task deep learning network for automatic calculation of left ventricular ejection fraction in 2d echocardiography. Computers in Biology and Medicine **156**, 106705 (2023)
11. Liang, M., Bian, B., Yang, Q.: Characteristics and long-term prognosis of patients with reduced, mid-range, and preserved ejection fraction: A systemic review and meta-analysis. Clinical Cardiology **45**(1), 5–17 (2022)
12. 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)
13. 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)
14. 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)
15. Pombo, J.F., Troy, B.L., RUSSELL JR, R.O.: Left ventricular volumes and ejection fraction by echocardiography. Circulation **43**(4), 480–490 (1971)
16. Reddy, C.D., Lopez, L., Ouyang, D., Zou, J.Y., He, B.: Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. Journal of the American Society of Echocardiography **36**(5), 482–489 (2023)
17. Rozenbaum, Z., Topilsky, Y., Khoury, S., Pereg, D., Laufer-Perl, M.: Association of body mass index and diastolic function in metabolically healthy obese with preserved ejection fraction. International Journal of Cardiology **277**, 147–152 (2019)
18. Savarese, G., Stolfo, D., Sinagra, G., Lund, L.H.: Heart failure with mid-range or mildly reduced ejection fraction. Nature Reviews Cardiology **19**(2), 100–116 (2022)
19. Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences. In: International conference on machine learning. pp. 3145–3153. PMLR (2017)
20. Steininger, M., Kobs, K., Davidson, P., Krause, A., Hotho, A.: Density-based weighting for imbalanced regression. Machine Learning **110**, 2187–2211 (2021)
21. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. arXiv preprint arXiv:2210.10163 (2022)
22. Weir, C.B., Jan, A.: Bmi classification percentile and cut off points (2019)
23. Williams, K., Thomson, D., Seto, I., Contopoulos-Ioannidis, D., Ioannidis, J., Curtis, S., Constantin, E., Batmanabane, G., Hartling, L., Klassen, T.: Standard 6: Age groups for pediatric trials. Pediatrics **129 Suppl 3**, S153–60 (06 2012). https://doi.org/10.1542/peds.2012-0055I
24. Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J.: Gnnexplainer: Generating explanations for graph neural networks. Advances in neural information processing systems **32** (2019)
25. Zhou, H.Y., Yu, Y., Wang, C., Zhang, S., Gao, Y., Pan, J., Shao, J., Lu, G., Zhang, K., Li, W.: A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. Nature Biomedical Engineering pp. 1–13 (2023)