ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096712

DEEP SPATIO-TEMPORAL MULTIPLEX GRAPH LEARNING FOR CARDIAC IMAGING CLASSIFICATION

Jaume Banus, Augustin Ogier*, Roger Hullin †, Philippe Meyer ‡, Ruud B. van Heeswijk*, Jonas Richiardi**

* Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland
† Department of Cardiology, Lausanne University Hospital and University of Lausanne, Switzerland
‡ Department of Medicine, Geneva University Hospital and University of Geneva, Switzerland

ABSTRACT

Cardiovascular diseases are the leading cause of mortality worldwide. Cardiac imaging is critical for precise characterization of cardiac structure and function, and is key in diagnosis, therapeutic management, and prognosis. In this work, we propose a novel representation of spatio-temporal cardiac data as a multiplex graph and develop a multi-level message passing neural network to classify clinical groups corresponding to different cardiovascular diseases. Using open data from the Automated Cardiac Diagnosis Challenge (N=150), our results show that the multiplex representation extracts discriminative features from the data (up to 94% accuracy compared to up to 80% for well-tuned baselines such as XGboost and MLP). Ablation studies show the bias induced by the graph's spatio-temporal structure improves generalization.

Index Terms— Message passing, Graph neural networks, Cardiac imaging, Multiplex graphs.

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of mortality globally, with an estimated prevalence of around 6'000/100'000 [1]. In this context, the analysis of the cardiac function is key for patient management, diagnosis and risk evaluation. Image-derived features are increasingly used in clinical routine. Most features, including stroke volume or ejection fraction (EF), are derived from automated segmentations obtained from algorithms such as convolutional neural networks (CNNs) [2]. Nevertheless, in the clinical environment there is a large data heterogeneity stemming from hardware and sequence differences, generating images with a variable resolution, contrast, and signal-to-noise ratio [3]. These differences impair generalization and result in variable performance when applied to unseen data. While data preprocessing and bias correction [4] are used to minimize these effects, processing pipelines themselves add another source of variability. Therefore, introducing domain knowledge as priors in those algorithms could help to mitigate measurement differences. Graph neural networks (GNNs) offer a possibility to include prior knowledge, e.g. by encoding the position of anatomical structures or functional connectivity [5, 6]. Their flexibility to handle structured and non-structured data with different relationships makes them a promising tool in medical diagnosis [7]. GNNs exploit spatial dependencies between entities under the main assumption that the state of a node relates to the state of its neighbors. Most of the work done in GNNs focused on message passing (MP) [8] and graph convolutional networks (GCNs) [9]. These approaches learn a latent feature representation on each node by aggregating information of a node and its neighbors, and proved to be an efficient solution for a variety of tasks including classification and regression at multiple hierarchical levels (nodes, edge and graph) and in different settings, such as supervised, unsupervised and semi-supervised learning [10]. Recently, GNNs have been extended to heterogeneous graphs, with multiple types of nodes and edges [11]. We propose to represent the heart as a multiplex graph whose nodes correspond to cardiac regions connected among them based on their anatomical location. Each node has a set of image-derived features, e.g. thickness or volume. Heterogeneous edges connect these nodes carrying information relative to the differences between cardiac regions. Moreover, cardiac images capture a temporal sequence that spans from end-diastole (ED) to end-systole (ES), adding a temporal dimension to the graph. Our goal is to use the multiplex graph representation to derive latent representations that capture the dynamics of the underlying system, in this case the heart contraction, and assess the influence of the added bias by the graph structure in the classification of clinical groups in the ACDC challenge [2] compared to classical approaches.

2. METHODS

2.1. Data processing and feature extraction

We used open data from the ACDC challenge [2], consisting of 150 patients divided into 5 evenly distributed subgroups: 1 normal control group (NOR) and 4 pathological groups which encompass myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and abnormal right ventricle (RV). The diagnostic criteria for each disease can be found in the challenge article [2]. The dataset was split in a balanced way in a training set of 100 sub-

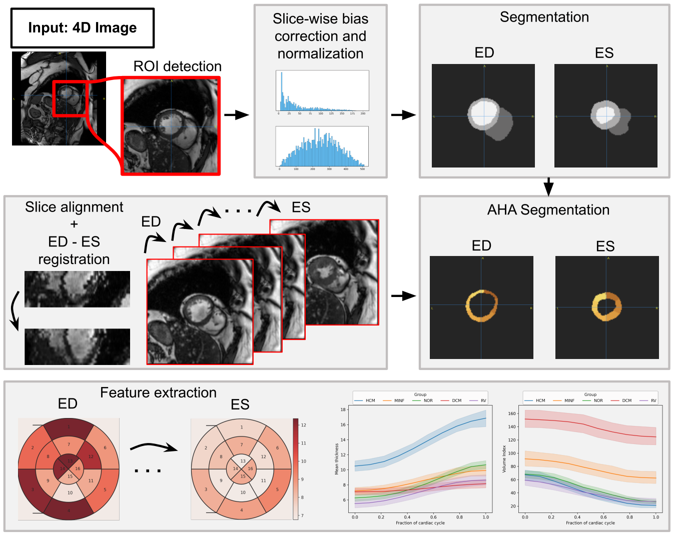


Fig. 1. Data processing pipeline. The input is a 4D image composed of 2D short-axis slices. We perform slice-wise bias-field correction and normalization, and use [12] to obtain the segmentation of the whole 4D volume. Next, we align the slices of the 4D volume to correct breath-motion artifacts and obtain the diffeomorphism from ED to ES. The LV myocardium at ED is divided in 16 AHA segments and the ED-ES diffemorphism is used to map the AHA labels at each time frame from ED to ES. For each segment we extract features associated to its thickness, volume and intensity distribution.

jects and a blind test set of 50 used to evaluate algorithms in the challenge. Each patient has a series of short-axis slices stacked in a 4D (x-y-z-t) volume covering the base to the apex, with a variable in-slice resolution around $1.5mm^2$ and a slice thickness from 5 to $10mm$. The entire dataset was acquired in clinical routine, resulting in an heterogeneous image quality. As illustrated in Figure 1, we processed the 4D images by performing slice-wise bias-field correction and normalization, and defined a region of interest (ROI) covering the whole heart. To segment the images we fine-tuned the model presented in [12] using the ED and ES ground-truth segmentation maps available in the training set. From the segmentation maps we computed the volumes of left ventricle (LV), right ventricle (RV), and the LV myocardium mass, as well as the ejection fraction (EF) for each region. We used the average center of mass of the LV at mid-cavity slices to re-align the 2D slices and correct the slice breath-motion misalignment. We divided the LV myocardium at ED in 16 segments according to the American Heart Association (AHA) model [13] and computed the volume and thickness of each segment. The apex region was not included given that for some subjects it is not in the field of view of the image. The division was done at ED since it represents the end of the filling phase, the instant at which the ventricles are expanded and filled with blood. From the thickness and intensity measurements of each region we extracted the mean, median, interquartile, lower quartile and upper quartile values. Following, we disassembled the 4D images and registered the consecutive frames from ES to ED using the symmetric normalization (SyN) registration algorithm, available in the advanced normalization tools (ANTs) package [14]. We obtained the diffeomorphism from ED to ES through the composition of the consecutive frames deformations. This allowed us to transfer the AHA labels to ES, and extract the volumes and thickness measurements at each frame up-to ES. Height and weight were used to compute the body surface area (BSA) according to Mosteller's formula [15]: $BSA = (\frac{Weight[kg]*Height[cm]}{3600})^{\frac{1}{2}}$, and used it to normalize the volumes of each region. Finally, we computed the Wasserstein distance between the distribution of voxel intensities of each region at each frame and between frames, as well as the distance between the centers of mass of each region. These distances are used as edge features in our multiplex graph to encode relative motion. Height, weight, BSA, and EF of the LV, RV and LV myocardium as well as stroke volume index were used as global features independent of the graph structure. Table 1 shows all features, their anatomical location, and whether they are node or edge attributes.

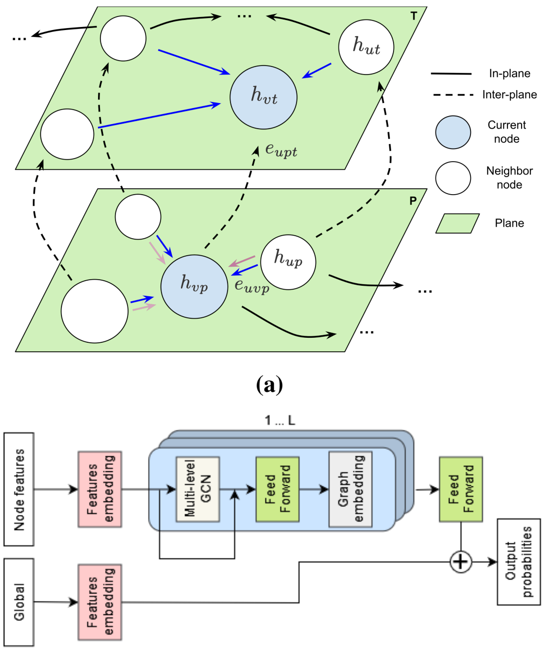
Table 1. Edge and node features. For the voxel intensity and thickness we extracted mean, median, lower quartile, upper quartile and interquartile values. Center of mass (CM), myocardium (Myo), left ventricle (LV), right ventricle (RV).

Feature	LV	RV	LV Myo	AHA	Type
Volume	✓	✓	✓	✓	Nodes
Voxel intensity	✓	✓	✓	✓	Nodes
Position (CM)	✓	✓	✓	✓	Nodes
Thickness			✓	✓	Nodes
Wasserstein distance	✓	✓	✓	✓	Edges
CM Distances	✓	✓	✓	✓	Edges

2.2. Graph representation

We represent the heart as a spatio-temporal multiplex (multilayer) graph, with the idea to exploit the interdependence between the different features and their spatio-temporal trajectory. In this context, each image frame can be viewed as a plane ($\mathcal{P}$) in the multiplex graph. In each plane there are multiple edges ($\mathcal{E}$) that connect the heart regions, mapped one-to-one to a set of vertices (or nodes) ($\mathcal{V}$). At the same time, the planes are connected between them through multiple edges ($\mathcal{T}$) encoding frame to frame differences in time. We formally represent the multiplex graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{P}, \mathcal{T})$. We considered as nodes the different AHA segments, as well as the LV myocardium, LV and RV blood pool regions. Typically, the adjacency matrix is based on a defined similarity between pairs of nodes. However, to include a spatial prior, we decided to connect each AHA segment to his adjacent segments. In turn, all the AHA segments are connected to the LV myocardium and to the RV and LV blood pools.

2.3. GNN implementation



(a)

(b)

Fig. 2. a) GCN layer blocks. Planes represent cardiac frames from the 4D image. Arrow colors indicate different edge features, dashed arrows show inter-plane edges, solid arrows show in-plane edges. b) End-to-end network architecture.

Our approach is mainly inspired by graph isomorphism networks (GIN) [16] and approaches based on MP neural networks [8]. In these approaches, the node representations are obtained by aggregating the neighbours' representations and combining them with the current node representation. Moreover, they can be extended to work with heterogeneous GNNs, as in [11] where each node process the messages of the different neighbors types independently and mixes them. In our case, we have a single node type, heart regions, but multiple messages. At each neural layer we have two messages that travel simultaneously through the graph in orthogonal directions, space and time. In particular, we have a spatial message (m_p) at each plane of the graph, and an additional temporal message (m_{pt}) that travels alongside the directed edges connecting consecutive graph planes (p and $t = p + 1$). This split of the MP introduces the notion of time in the network. The input to our layer (l) is a set of node features for each plane, $h^{l-1}_p = (h_1, h_2, ..., h_{\mathcal{V}_p}), h^{l-1}_{ip} \in \mathbb{R}^D$, where D is the number of features and $\mathcal{V}_p$ the number of nodes at the plane p. The layer produces a new set of node features per plane ($D^{'}$), $h^{'}_p = (h^{'}_1, h^{'}_2, ..., h^{'}_{\mathcal{V}_p}), h^{'}_{ip} \in \mathbb{R}^{D^{'}}$. The set of edges at each plane is denoted $\mathcal{E}_{pt} = (e_1, e_2, ...e_{E_{pt}}), e_i \in \mathbb{R}^{F_e}$, where F_e is the number of edge features and E_{pt} indicates the number of edges in the set, the set $\mathcal{E}_{pp}$ represents the in-plane edges. At layer l, the message at node v and plane p is defined as:

$$m^l_{vp} = \max_{u\in\mathcal{N}_p(v)} \left(\sum_f^{F_e} \frac{h^l_{up}\cdot e^f_{uvpp}}{c_{uvpp}} \right) + m^l_{vpt}, \qquad (1)$$

where m^l_{vpt} is defined as

$$m^l_{vpt} = \max_{u\in\mathcal{N}_{pt}(v)} \left(\sum_f^{F_e} \frac{\left(h^l_{ut} - h^l_{vp}\right)\cdot e^f_{uvpt}}{c_{uvpt}} \right), \qquad (2)$$

where $\mathcal{N}_p(v)$ denotes the neighbourhood of node v at plane p, c_{uvpt} represents the normalization coefficient based on the in-node degree, and f iterates over the edge features. The received messages are combined to update the node representation as $h^l_{vp} = \phi^l\left(\left(1+\epsilon\right)\cdot h^{l-1}_{vp} + m^l_{vp}\right)$, where ϕ^l is a 2-layer MLP with H-dimensional input and H-dimensional output and ϵ is a learnable parameter. For each neural network layer (l), a graph embedding is obtained based on the node embeddings as $g^l = \psi_G\left(\psi^l\left(h^l\right)\right)$, where ψ^l is parameterized as a 2-layer MLP with a H-dimensional output and a $H \cdot N$-dimensional input input, and ψ_G is a 2-layer MLP shared across the neural network layers with H-dimensional input and 5-dimensional output. The final graph embedding is obtained as the addition of the average latent representations of the total number of neural network layers (L) and the global features embedding (h_{ext}), as $G_h = \frac{1}{L}\left(\sum_{l=1}^{L} g^l\right) + h_{ext}$; where h_{ext} is obtained using a 2-layer MLP with H_g hidden dimensions and 5-dimensional output. Initial node features and global features were z-score normalised with respect to the healthy group. The graph embedding, G_h, is used to classify each graph into the 5 different groups using a softmax non-linearity to optimize a multi-class cross-entropy loss. The GNN was implemented using deep graph library (DGL) [17] with Pytorch [18] backend and optimized with Adam [19].

3. EXPERIMENTS AND RESULTS

We compared our method to a fully connected MLP, gradient-boosted decision trees (XGBoost), a random forest and a linear support vector machine (SVM). To obtain the best set of hyperparameters for each algorithm we performed nested cross-validation (CV) with 2 outer folds and 3 inner folds. At each outer fold we split the training set (N=100) of the ACDC challenge into 85:15 for training and test, and performed CV in the training set. For our model and the MLP the hyperparameter space was explored using optuna [20]. For the other methods, we relied on the HalvingGridSearch approach implemented in scikit-learn [21]. The set of selected hyperparameters for each classifier is available at Table 3. Finally, we split the training set in 85:15 and used the average of the two best set of hyperparameters to train the final classifier. Additionally, we performed an ablation study to assess the influence of the edge features and the graph spatial and temporal connectivity. Table 2 reports the mean and standard deviation of the accuracy obtained in CV and the results on the blind test set (N=50).

Table 2. Cross-validation (CV) and blind test set accuracy.

Method	CV Acc. (± Std)	Acc. Test	Edges	Global	Spatial	Temporal
Linear SVM	0.72 ± 0.091	0.76	✓	✓	✓	✓
Random Forest	0.80 ± 0.066	0.80	✓	✓	✓	✓
XGBoost	0.82 ± 0.049	0.84	✓	✓	✓	✓
MLP	0.85 ± 0.033	0.74	✓	✓	✓	✓
Ours and ablation tests						
Multiplex	0.93 ± 0.044	0.88	✓	✓	✓	✓
	0.91 ± 0.022	0.86	✓	✓	✓	✗
	0.90 ± 0.100	0.74	✓	✓	✗	✓
	0.90 ± 0.056	0.94	✓	✗	✓	✓
	0.85 ± 0.056	0.80	✗	✓	✗	✗
	0.80 ± 0.067	0.66	✗	✗	✗	✗

Table 3. Hyperparameter ranges and final value. The multiplex GNN and the MLP hyperparameter space were explored using Optuna, others with HalvingGridSearch.

Method	**Hyperparameter**	**Range**	**Value**
Multiplex	Learning rate	[1e-4 1e-2]	0.0038
	Batch size	[20 40]	24
	Num. Layers	[2 6]	3
	Hidden nodes	[15 80]	34
	Hidden global	[15 50]	45
	Dropout rate	[0.1 0.4]	0.317
	L2 weight	[0 0.3]	0.09
	L1 weight	[0 0.3]	0.15
MLP	Learning rate	[1e-5 1e-2]	0.00028
	Batch size	[20 30]	21
	Num. layers	[2 10]	5
	Hidden size	[50 250]	190
	Dropout rate	[0 0.4]	0.15
	L2 weight	[0 0.2]	0.10
	L1 weight	[0 0.2]	0.19
Random Forest	Estimators	[501 601]	601
	Criterion	[Gini Entropy]	Gini
	Class weight	[Balanced Balanced(subsample)]	Balanced(subsample)
XGBoost	Estimators	[501 601]	601
	Alpha	[0 1]	0
	Lambda	[0 1]	0
	Subsample	[0 1]	0.33
Linear SVM	C	[1e-8 1e-3]	1e-3
	Class weight	[None Balanced]	Balanced

4. DISCUSSION

Results presented in Table 2 indicate that our approach outperforms classical methods used to deal with tabular data even when using fewer features. Regarding the ablation tests, we obtained consistently worse results when edges were set to 0. In turn, in-plane spatial connectivity seems to be more important than inter-plane connections. Nevertheless, both temporal and spatial connectivity are needed to achieve the best performance. Moreover, the results suggest that edges have a higher importance than graph external data. Overall, these results highlight that structural information present in the graph helps to generalize to unseen data. To the best of our knowledge, this work is the first to represent the cardiac cycle as a spatio-temporal multiplex graph. Previous work [22] represented the LV myocardium as point cloud data and used a spatio-temporal GCN [23] to predict the contour position in future frames. However, in the ST-GCN setting, all nodes are connected and edge features are not considered. Moreover, we are using the temporal edge to compute a delta, similar to a discrete step in an Eulerian discretization. The use of directed edges in time encodes the sequence temporal ordering. Therefore, we introduced a novel formulation based on MP that allow us to include directionality in the data and implicitly represent time. Recent works proposed more expressive architectures than MP, mainly related to higher order GNNs that allow the identification of graph substructures [24, 25]; we note that these approaches can still be viewed from a MP point of view [26]. Nevertheless, they could be of interest for the generalization to non-parcellated data, e.g. we could use mesh-like data without a one-to-one node correspondence between the mesh nodes of different subjects. Another point of interest is the generation of the adjacency matrix. It would be interesting to explore the simultaneous learning of the adjacency matrix during task optimization [27] and use the AHA connectivity pattern as prior knowledge to regularize the structural learning. Regarding the temporal aspect, most of the approaches are sequence-to-sequence methods focused on forecast multivariate time-series. These methods are usually based on recurrent neural networks [27], 1D convolutions [28], or transformer-like architectures [29]. Currently, we use the temporal aspect to add a direction in our data and exploit the relative differences between the node states in time. Nevertheless, a sequence-to-sequence approach could be implemented and use the error in the frame prediction as an unsupervised metric to identify the most meaningful meta-paths in the heterogeneous multiplex, similar to the approach present in [30] to detect anomalies in cardiac shapes from point cloud data.

5. CONCLUSIONS

To the best of our knowledge this work represents the first attempt to model the cardiac dynamics from an heterogeneous spatio-temporal graph perspective, and we showed that our message-passing GNN is able to discriminate between different cardiac diseases with an accuracy comparable to the top three algorithms in the ACDC challenge which achieved an accuracy in the blind test set of at least 0.92. In future work, we want to assess the robustness of our approach to uncertainty by applying the learned model to other datasets, such as UK Biobank [31]. Moreover, we will explore end-to-end learning of the graph structure [27] and the interpretability of our results. The spatio-temporal graph-based structure opens the door to identifying which cardiac regions are more important and when for different diseases, which would be a valuable tool for clinical diagnosis.

6. ACKNOWLEDGEMENT

This research was supported supported by the Swiss National Foundation Sinergia Grant CRSII5_202276/1.

References

[1] Gregory A. Roth et al., "Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019," *Journal of the American College of Cardiology*, vol. 76, Dec. 2020.

[2] Olivier Bernard et al., "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, 2018.

[3] Frank J Brooks and Perry W Grigsby, "Quantification of heterogeneity observed in medical images," *Medical Imaging*, vol. 13, pp. 7, 2013.

[4] Kerstin N. Vokinger and et al., "Mitigating bias in machine learning for medicine," *Communications Medicine*, vol. 1, 12 2021.

[5] Karthik Gopinath and et al., "Graph domain adaptation for alignment-invariant brain surface segmentation," *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis : UNSURE and GRAIL*, 2020.

[6] Sarah Parisot and et al., "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease," *Medical Image Analysis*, vol. 48, pp. 117–130, 8 2018.

[7] David Ahmedt-Aristizabal and et al., "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," *Sensors*, vol. 21, pp. 4758, 7 2021.

[8] Justin Gilmer and et al., "Neural message passing for quantum chemistry," *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[9] D. I. Shuman and et al., "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, pp. 83–98, 5 2013.

[10] Zonghan Wu and et al., "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, 2021.

[11] Chuxu Zhang and et al., "Heterogeneous graph neural network," *Proceedings of the 25th International Conference on Knowledge Discovery & Data Mining*, 2019.

[12] Wenjia Bai and et al., "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *JCMR*, vol. 20, 9 2018.

[13] Manuel D. Cerqueira and et al., "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart," *Circulation*, vol. 105, 1 2002.

[14] B. B. Avants and et al., "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[15] RD Mosteller, "Simplified calculation of body-surface area," *The new england journal of medicine*, 1987.

[16] Keyulu Xu and et al., "How powerful are graph neural networks?," *International Conference on Learning Representations*, 2019.

[17] Minjie Wang and et al., "Deep graph library: A graph-centric, highly-performant package for graph neural networks," *'Arxiv'*, 9 2019.

[18] Adam Paszke and et al., "Pytorch: An imperative style, high-performance deep learning library," *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.

[19] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *3rd ICLR*, 2015.

[20] Takuya Akiba and et al., "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[21] Fabian Pedregosa and et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[22] Ping Lu and et al., "Modelling cardiac motion via spatio-temporal graph convolutional networks to boost the diagnosis of heart conditions," *STACOM*, 2021.

[23] Sijie Yan et al., "Spatial temporal graph convolutional networks for skeleton-based action recognition," *Proceedings of the Thirty-Second AAAI*, 2018.

[24] Cristian Bodnar and et al., "Weisfeiler and lehman go cellular: CW networks," *Advances in Neural Information Processing Systems*, 2021.

[25] Erik Henning Thiede and et al., "Autobahn: Automorphism-based graph neural nets," *Advances in Neural Information Processing Systems*, 2021.

[26] P. Veličković, "Message passing all the way up," 2022.

[27] Chao Shang and et al., "Discrete graph structure learning for forecasting multiple time series," *International Conference on Learning Representations*, 2021.

[28] Zonghan Wu and et al., "Connecting the dots: Multivariate time series forecasting with graph neural networks," *Proceedings of the 26th International Conference on Knowledge Discovery & Data Mining*, 2020.

[29] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun, "Heterogeneous graph transformer," *Proceedings of The Web Conference 2020*, p. 2704–2710, 2020.

[30] Arezoo Zakeri and et al., "A probabilistic deep motion model for unsupervised cardiac shape anomaly assessment," *Medical Image Analysis*, vol. 75, 1 2022.

[31] C. Sudlow and et al., "UK Biobank an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.," *PLoS Med.*, 2015.