Cross-Modality Cardiac Insight Transfer: A Contrastive Learning Approach to Enrich ECG with CMR Features

Zhengyao Ding[1], Yujian Hu[2], Ziyu Li[1], Hongkun Zhang[2], Fei Wu[1], Yilang Xiang[2], Tian Li[3], Ziyi Liu[4], Xuesen Chu[5(✉)], and Zhengxing Huang[1]

[1] Zhejiang University, China
{zhengyao.ding,liziyu,wufei,zhengxinghuang}@zju.edu.cn
[2] The First Affiliated Hospital of Zhejiang University School of Medicine, China
{3170103999,1198050,21618130}@zju.edu.cn
[3] The Hong Kong Polytechnic University tianli@polyu.edu.hk
[4] Guangdong Transtek Medical Electronics Co., Ltd., Zhongshan, 528437, China
11313008@zju.edu.cn
[5] China ship scientific research center chuxs@cssrc.com.cn

Abstract. Cardiovascular diseases are the leading cause of death worldwide, and accurate diagnostic tools are crucial for their early detection and treatment. Electrocardiograms (ECG) offer a non-invasive and widely accessible diagnostic method. Despite their convenience, they are limited in providing in-depth cardiovascular information. On the other hand, Cardiac Magnetic Resonance Imaging (CMR) can reveal detailed structural and functional heart information; however, it is costly and not widely accessible. This study aims to bridge this gap through a contrastive learning framework that deeply integrates ECG data with insights from CMR, allowing the extraction of cardiovascular information solely from ECG. We developed an innovative contrastive learning algorithm trained on a large-scale paired ECG and CMR dataset, enabling ECG data to map onto the feature space of CMR data. Experimental results demonstrate that our method significantly improves the accuracy of cardiovascular disease diagnosis using only ECG data. Furthermore, our approach enhances the correlation coefficient for predicting cardiac traits from ECG, revealing potential connections between ECG and CMR. This study not only proves the effectiveness of contrastive learning in cross-modal medical image analysis but also offers a low-cost, efficient way to leverage existing ECG equipment for a deeper understanding of cardiovascular health conditions. Our code is available at https://github.com/Yukui-1999/ECCL.

Keywords: Contrastive Learning · Cross-Modal Medical Image Analysis · ECG-CMR Integration.

First Author and Second Author contribute equally to this work.

1 Introduction

Cardiovascular diseases are the foremost cause of mortality globally, posing a significant threat to human health [18]. Early detection and treatment of these diseases are imperative, necessitating the reliance on precise diagnostic tools. Among the various diagnostic methods, Electrocardiograms (ECG) emerge as the preferred choice due to their non-invasive nature, simplicity of operation, and widespread availability [27]. ECGs are capable of revealing basic cardiac-related features such as heart rate and arrhythmias, providing essential diagnostic information for the preliminary detection of cardiac anomalies. In recent years, ECG-based deep learning models have been increasingly applied in the analysis of cardiovascular diseases, showcasing the significant potential of ECG in detecting and classifying various cardiovascular conditions [3,16,25]. These models employ advanced machine learning techniques [22,28,10] and leverage large datasets of ECG recordings, allowing for the identification of subtle patterns and abnormalities that may be indicative of cardiovascular-related diseases [1,12,24,9,13,2]. Despite the significant role that ECG plays in routine clinical diagnostics, its capacity to provide detailed cardiovascular information is limited. In contrast, Cardiac Magnetic Resonance Imaging (CMR) offers comprehensive phenotypic and morphological descriptions of the heart, including advanced information on cardiac structure, function, and tissue characteristics [26], establishing it as the gold standard for evidence-based diagnosis of various cardiovascular diseases [14]. However, the complexity of CMR operations, its high cost, and the technical expertise required for operators restrict its use in primary healthcare institutions, particularly in rural hospitals [11].

Considering the limitations of ECG in delivering comprehensive cardiovascular insights, the emergence of multi-modal contrastive learning, inspired by foundational projects like ConVIRT [29] and CLIP [20], offers a promising pathway to augment ECG analysis. Notably, research efforts such as those by Qiu et al [19]. and Liu et al [15]. have pioneered the use of contrastive learning between ECG signals and textual data to bolster model efficacy for downstream applications. Furthermore, the innovative proposal by Radhakrishnan et al [21]. for a cross-modal autoencoder that bridges ECG and CMR technologies aims to furnish a thorough cardiovascular health assessment utilizing solely ECG data. This approach, however, raises concerns regarding potential inconsistencies in the encoding of data across different training phases, which could hinder the achievement of a cohesive representation and impede the full transfer of information. To surmount these challenges and effectively close the diagnostic divide between ECG and CMR for cardiovascular disease identification, we introduce a novel contrastive learning methodology. This method strategically freezes the CMR encoder after pre-training, facilitating a seamless and efficient transfer of data from CMR to ECG within the latent space. This innovation significantly augments the diagnostic utility of ECG-based models in identifying and classifying cardiovascular diseases. Our primary contributions are as follows:

– We propose a contrastive learning framework to transfer detailed cardiac information from CMR into the feature embeddings of ECG, overcoming the limitations of ECG in providing deep cardiovascular information.
– We validate our method on 41,519 samples from the UK Biobank (UKB). The achieved results demonstrate that our method significantly enhances the predictive capability of ECG for cardiac phenotype prediction, as well as for the prediction of conditions such as myocardial infarction and heart failure.
– We conduct extensive ablation experiments to verify the effectiveness of each component of our model, ensuring a comprehensive understanding of its contributions to the overall performance.

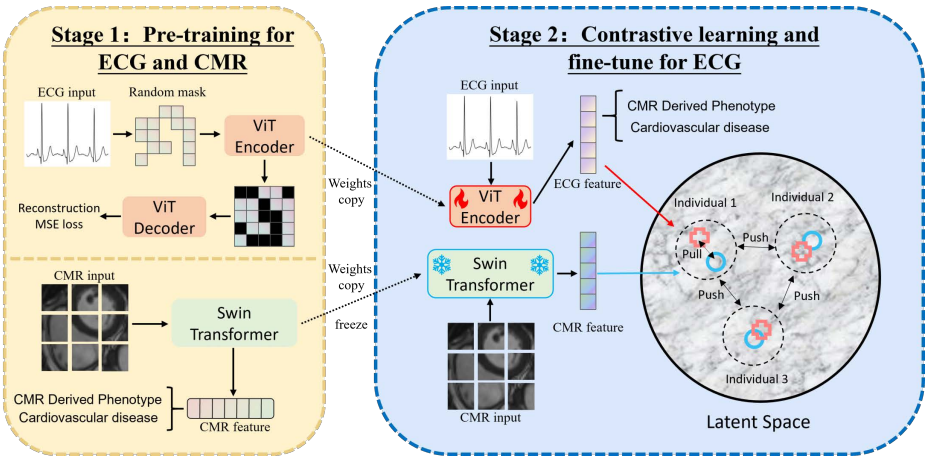

Fig. 1. ECCL is structured into two stages. In the first stage, we conduct self-supervised training for ECG and supervised training for CMR. In the second stage, we utilize the CMR encoder trained during the first stage and proceed with contrastive learning for both ECG and CMR, along with supervised fine-tuning for ECG.

2 Method

We propose a novel cross-modal contrastive learning approach, dubbed ECCL (Electrocardiogram-Cardiac Contrastive Learning), that utilizes deep learning techniques to explore the relationship between ECG and CMR data. Our proposed ECCL aims to enhance the overall representation capability of ECG as a single modality for cardiovascular analysis. As shown in Fig. 1, our method involves a first phase of self-supervised pre-training for ECG and supervised pre-training for CMR, followed by a second phase of cross-modal contrastive learning and supervised fine-tuning for ECG.

2.1 Self-supervised Pre-training for ECG and Supervised Pre-training for CMR

We utilize ViT [6] and Swin Transformer [17] as encoders for ECG and CMR, respectively. For the self-supervised pre-training of ECG, we employ the Masked Autoencoder (MAE) architecture as outlined by He et al. [7]. By selectively masking portions of the input data and then tasking the model with predicting these masked values, the MAE framework encourages the learning of more comprehensive and nuanced embeddings. This approach not only facilitates the encoder in capturing a richer representation of the underlying ECG signals but also enhances its ability to discern subtle patterns and variations within the data. Considering an ECG signal with dimensions $x \in \mathbb{R}^{C \times T}$, where C represents the number of leads and T represents the number of sampling points, we utilize the ViT methodology to partition the $C \times T$ signal into multiple tokens, thus transforming the data format into $L \times D$, where L is the number of tokens and D is the dimension size of each token. Subsequently, we randomly mask L_m tokens, leaving L_v tokens visible, where $L = L_m + L_v$. Let f be the MAE encoder and g the MAE decoder, then $\hat{x} = g(f(X_v))$. By minimizing the Mean Squared Error (MSE) loss of the reconstruction, we can obtain an ECG encoder capable of generating rich embedding information.

For CMR data, we opt for a supervised learning approach during pre-training. To fully explore the information potential within CMR, we utilize the 82 cardiac phenotype indicators found in the UK Biobank dataset, including LV myocardial mass, and LV end-diastolic volume, among others, with a complete list available in the supplementary materials [4]. Additionally, we conduct training for specific conditions like myocardial infarction (ICD10 I21), cardiomyopathy (ICD10 I42), atrial fibrillation (ICD10 I48), and heart failure (ICD10 I50). When dealing with cardiac phenotypes, we use Mean Squared Error as the loss function for regression training, while for diseases, we employ Binary Cross-Entropy Loss for classification training. Given the low prevalence of disease cases, typically around 2% of the dataset, we implement a strategy where we combine all positive cases with a randomly selected subset of negative samples to create a small sample set for iterative training. Through this supervised training process, we develop five pre-trained CMR encoders, which will be frozen in subsequent steps to facilitate cross-modal contrastive learning with ECG.

2.2 Cross-Modal Alignment with Frozen CMR Encoder and Supervised Training for ECG

During the cross-modal alignment stage, we employed a loss function similar to CLIP to achieve the alignment of ECG and CMR in the latent space. Assuming the ECG encoder as f_e, which was copied weight from the encoder after SSL in stage one, the projection head as g_e, the CMR encoder as f_c, the projection head as g_c, with x_e and x_c representing the input ECG and CMR data respectively, then the features in the latent space for ECG are $z_e = g_e(f_e(x_e))$, and for CMR

are $z_c = g_c(f_c(x_c))$. The loss for ECG is defined as:

$$\mathcal{L}_{\text{ecg}} = -\log \frac{\exp\left(z_e^\top z_c/\tau\right)}{\sum_{j=0}^{k} \exp\left(z_e^\top z_c^j/\tau\right)} \tag{1}$$

The loss for CMR is defined as:

$$\mathcal{L}_{\text{cmr}} = -\log \frac{\exp\left(z_c^\top z_e/\tau\right)}{\sum_{j=0}^{k} \exp\left(z_c^\top z_e^j/\tau\right)} \tag{2}$$

The total contrastive learning loss is:

$$\mathcal{L}_{\text{CL}} = \frac{1}{2}\mathcal{L}_{\text{ecg}} + \frac{1}{2}\mathcal{L}_{\text{cmr}} \tag{3}$$

To maintain the stability of the CMR encoder during the subsequent steps, we froze the CMR encoder during the cross-modal alignment training phase. By doing so, we preserve the feature representations learned during the supervised pre-training stage. Freezing the parameters of the CMR encoder ensures that the representations on the CMR side remain stable and consistent when aligning with ECG data. This is essential for preserving the quality of information transfer between the modalities.

In addition to contrastive learning, we conduct supervised training on ECG data for specific tasks. We use the same labels and cardiac phenotype indicators as employed in the supervised training phase for CMR. These indicators cover four categories of cardiovascular diseases—heart failure, myocardial infarction, cardiomyopathy, and atrial fibrillation—available in the UK Biobank (UKB) dataset. Assuming the classification head as h, and the model loss as:

$$\mathcal{L}_{\text{task}} = loss\_fn(h(f_e(x_e)), label) \tag{4}$$

we employ MSE Loss as $loss\_fn$ for regression of cardiac phenotype indicators and BCE Loss as $loss\_fn$ for disease classification. Given the scarcity of positive samples, we adopt an iterative training approach that combines all positive and a randomly selected subset of negative samples. Thus, the total loss for the second phase is formulated as:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{CL}} + \lambda\mathcal{L}_{\text{task}} \tag{5}$$

where λ is a weighting coefficient used to balance the contrastive learning loss $\mathcal{L}_{CL}$ and the task-specific loss $\mathcal{L}_{task}$.

This approach not only achieves effective alignment between ECG and CMR data in the latent space but also enhances the accuracy and robustness of the ECG encoder in downstream cardiovascular disease diagnostic tasks.

3 Experiments

3.1 Dataset and experimental details

This study is based on data provided by the UK Biobank (UKB). We utilized data from the first imaging assessment in UKB, focusing on electrocardiogram

(ECG) data with a standard sampling frequency of 500Hz and a sampling duration of 10 seconds. To address the issue of baseline drift in ECG signals, we employed the seasonal decompose method for preprocessing. For cardiac magnetic resonance (CMR) imaging, we selected 50 frames of images from the middle basoapical slice. Following the method described in literature [5], we segmented the CMR images and cropped the smallest bounding rectangle containing the heart, then resized it to 224×224 pixels. Consequently, the input data for the CMR images was organized into a matrix of 50 channels of 224×224 pixels. The entire dataset comprises 41,519 samples, with 24,908 allocated for the training set, 8,303 for the validation set, and 8,308 for the test set.

In our experiments, we employed the PyTorch framework version 2.1.2 for model training and testing. The processing of ECG data was based on the basic Vision Transformer (ViT) model, setting the patch size to 1x100, with the dimension of the model's embedding layer at 768, including 12 Transformer layers, each with 12 attention heads. For CMR image processing, we selected the Swin Transformer model, with a patch size set to 4x4 and a window size of 7. At the start of training, we applied data augmentation techniques, including random cropping, temporal flipping, and spatial flipping for ECG data, and random rotation and scaling for CMR data, normalizing them to the range of [-1,1]. During the self-supervised training phase for ECG data, we set a masking ratio of 0.8. For contrastive learning, ECG and CMR data were independently encoded into a shared feature space with 256 dimensions and set the λ in the total loss function to 1. In the classification task, given the scarcity of positive samples, with prevalence rates of 2.2% for Myocardial Infarction (I21), 0.4% for Cardiomyopathy (I42), 3.9% for Atrial Fibrillation (I48), and 1.2% for Heart Failure (I50), we employed a strategy of combining all positive samples with a subset of negative samples (twice the number of positive samples) to form several sub-datasets. Iterative training was conducted within these sub-datasets. The Adam optimizer was used for model training, with a learning rate of 8e-4 for ECG self-supervised training and 8e-5 for the remaining training phases. We adjusted the learning rate using a cosine annealing algorithm, with a 40-round warm-up period and a total of 200 training rounds planned. All model parameter adjustments and training processes were completed on a 24 GB NVIDIA GeForce RTX 4090 GPU.

3.2 Results

We leveraged ECG data enhanced through contrastive learning for predicting 82 cardiac phenotype indicators (LV/RV end-diastolic volume, LV/RV end-systolic volume, LV/RV stroke volume, and so on) and diagnosing heart diseases (myocardial infarction: ICD code I21, cardiomyopathy: ICD code I42, atrial fibrillation: ICD code I48, and heart failure: ICD code I50) to evaluate our model.

Baselines: Given the scarcity of research employing contrastive learning with ECG and CMR data, we compared following methods: CMAE [21], the modifications made to our CMR encoder(ResNet50 [8] and ViT [6]), and a triplet loss approach [23]. **Evaluation Metrics**: We employed the Pearson correlation

coefficient to evaluate the performance of our model in cardiac phenotype regression tasks. For classification tasks (diagnosing heart diseases), the Area Under the Curve (AUC) metric was used as the evaluation criterion. **Ablation Study:** We compared ECG signals under three conditions: neither self-supervised nor contrastive learning, self-supervised but not contrastive learning, and not self-supervised but contrastive learning. Additionally, we provided results solely using CMR data supervised training based on the Swin Transformer as a reference. All results are presented in Table 1, with five different random seeds. We also did experiments comparing different ECG patch numbers and ViT sizes in the regression task, with results presented in Fig.2(b).

The results demonstrate that our approach significantly improves the ability of ECG to detect cardiomyopathy and myocardial infarction. Remarkably, for cardiomyopathy, our method achieves even better results on ECG than supervised CMR, leading us to infer that for certain cardiovascular diseases, ECG and CMR may contain some mutually exclusive features. Only through multimodal learning can these features be unlocked, thereby enhancing the guidance for diagnosis and analysis of diseases. However, the results of using ECG to regress the 82 cardiac phenotype indicators still significantly lag behind those obtained from supervised CMR, which will be the focus of our future efforts. Moreover, it is observed in Fig. 2 that the number of ECG patches significantly affects ECG representation, while increasing the size of ViT can improve performance metrics but is not cost-effective in terms of the additional GPU memory and training time required. Additionally, we utilized the T-SNE to visualize the distribution of ECG and CMR in the latent space before and after contrastive learning, as illustrated in Fig. 2(a). Also, we used UMAP for disease-specific visualization of ECG features before and after alignment, as shown in Fig. 2(d). It is evident that the aligned ECG features show clearer separation between positive and negative samples. UMAP results for all diseases are available in the supplementary materials. Fig. 2(c) presents the partial results of regression analysis on 82 cardiac indicators within our test set, comprising 8308 samples in total. For more detailed outcomes, please refer to the appendix.

Table 1. Comparison of previous methods and our proposed ECCL and ablation experiments. MI, CM, AF, and HF stand for diseases and Mean R stands for the regression correlation coefficient of the cardiac phenotype prediction.

	MI AUC ↑	CM AUC ↑	AF AUC ↑	HF AUC ↑	Mean R ↑
CMAE [21]	0.705±0.005	0.733±0.058	0.739±0.002	**0.818±0.006**	0.387±0.001
ResNet50 [8]	0.703±0.004	0.742±0.008	0.715±0.012	0.768±0.009	0.382±0.003
ViT [6]	0.710±0.012	0.775±0.017	0.708±0.007	0.793±0.004	0.395±0.002
Triplet [23]	**0.725±0.004**	0.745±0.003	0.730±0.008	0.795±0.010	0.372±0.005
W/O SSL; W/O CL	0.627±0.007	0.565±0.027	0.634±0.001	0.664±0.021	0.252±0.012
W/ SSL; W/O CL	0.702±0.002	0.794±0.014	0.737±0.005	0.787±0.005	0.351±0.003
W/O SSL; W/ CL	0.683±0.010	0.612±0.009	0.698±0.003	0.714±0.015	0.339±0.002
ECCL(Ours)	0.714±0.009	**0.826±0.012**	**0.739±0.001**	0.807±0.005	**0.407±0.002**
CMR_Sup	0.729±0.006	0.740±0.052	0.767±0.003	0.761±0.014	0.561±0.007

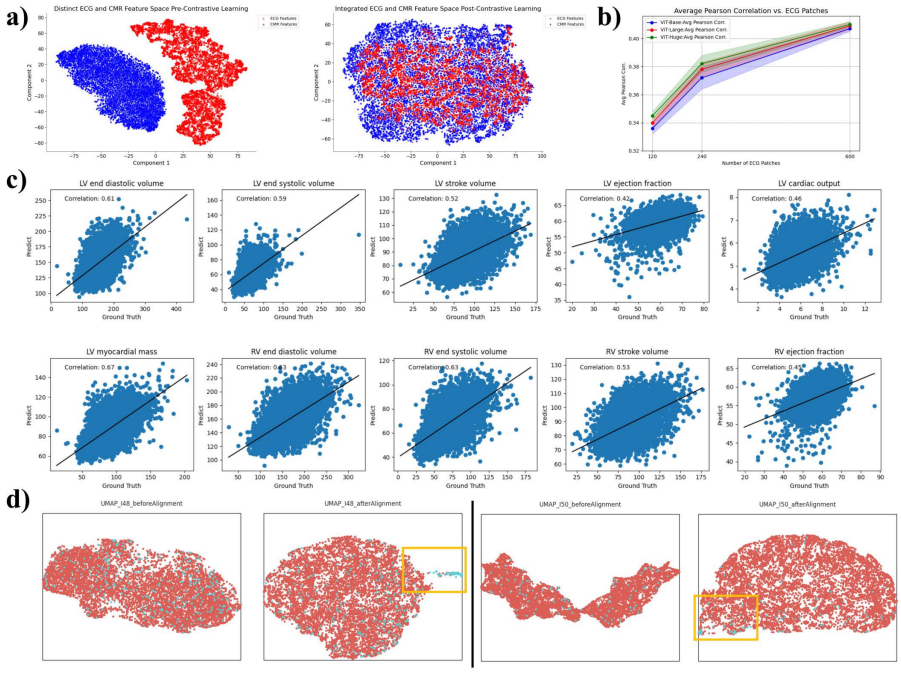

Fig. 2. Visualization of results: (a) T-SNE results of distribution for ECG and CMR in the hidden space before (left panel) and after (right panel) comparative learning. (b) Mean Pearson correlation coefficients in regression tasks using different numbers of ECG Patches and ViT sizes. (c) Results of Pearson's correlation coefficients for some of the cardiac trait metrics in the regression task, and the whole results of 82 cardiac traits can be found in the supplementary material. (d) UMAP visualization of ECG features before and after alignment for certain diseases, with blue representing positive samples and red representing negative samples.

4 Conclusion and Discussion

This study introduces a novel contrastive learning framework that significantly boosts the diagnostic accuracy of ECG by incorporating insights from CMR, effectively closing a vital gap in non-invasive cardiovascular diagnostics. By integrating CMR data through a contrastive learning approach, our findings showcase substantial enhancements in diagnosing cardiovascular diseases and predicting cardiac phenotypes solely using ECG data. These advancements represent a leap forward in our ability to leverage ECG for detailed cardiovascular analysis. However, despite the progress, challenges persist in fully encapsulating the comprehensive spectrum of cardiac traits via ECG, which may become a direction for subsequent research and a focal point for breakthroughs.

This research not only confirms the immense potential of cross-modal data integration in the realm of medical imaging but also underscores the extensive utility of ECG as an affordable, yet powerful, tool for a thorough assessment of cardiovascular health. This innovative approach aims to broaden the diagnostic capabilities of ECG, making it a more potent tool in the fight against cardiovascular diseases by enriching it with the depth of information typically reserved for more invasive diagnostic methods.

Acknowledgments. This work was supported by the Technical Innovation key project of Zhejiang Province (2024C03023) to H.Z, the National Key Research and Development Program of China (Grant No. 2022YFF1202400), and the National Nature Science Foundation of China (Grant No. 82272129). The authors thank the UK Biobank (UKB) for providing data under Application ID 89757.

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

References

1. Akbilgic, O., Butler, L., Karabayir, I., Chang, P.P., Kitzman, D.W., Alonso, A., Chen, L.Y., Soliman, E.Z.: Ecg-ai: electrocardiographic artificial intelligence model for prediction of heart failure. European Heart Journal-Digital Health **2**(4), 626–634 (2021)
2. Al-Zaiti, S.S., Martin-Gill, C., Zègre-Hemsey, J.K., Bouzid, Z., Faramand, Z., Alrawashdeh, M.O., Gregg, R.E., Helman, S., Riek, N.T., Kraevsky-Phillips, K., et al.: Machine learning for ecg diagnosis and risk stratification of occlusion myocardial infarction. Nature Medicine **29**(7), 1804–1813 (2023)
3. Attia, Z.I., Harmon, D.M., Behr, E.R., Friedman, P.A.: Application of artificial intelligence to the electrocardiogram. European heart journal **42**(46), 4717–4730 (2021)
4. Bai, W., Suzuki, H., Huang, J., Francis, C., Wang, S., Tarroni, G., Guitton, F., Aung, N., Fung, K., Petersen, S.E., et al.: A population-based phenome-wide association study of cardiac and aortic structure and function. Nature medicine **26**(10), 1654–1662 (2020)
5. Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P.M., Rueckert, D.: Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11. pp. 586–594. Springer (2018)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16000–16009 (2022)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

9. Khurshid, S., Friedman, S., Reeder, C., Di Achille, P., Diamant, N., Singh, P., Harrington, L.X., Wang, X., Al-Alusi, M.A., Sarma, G., et al.: Ecg-based deep learning and clinical risk factors to predict atrial fibrillation. Circulation **145**(2), 122–133 (2022)
10. Kiyasseh, D., Zhu, T., Clifton, D.A.: Clocs: Contrastive learning of cardiac signals across space, time, and patients. In: International Conference on Machine Learning. pp. 5606–5615. PMLR (2021)
11. von Knobelsdorff-Brenkenhoff, F., Pilz, G., Schulz-Menger, J.: Representation of cardiovascular magnetic resonance in the aha/acc guidelines. Journal of Cardiovascular Magnetic Resonance **19**(1), 70 (2016)
12. Ko, W.Y., Siontis, K.C., Attia, Z.I., Carter, R.E., Kapa, S., Ommen, S.R., Demuth, S.J., Ackerman, M.J., Gersh, B.J., Arruda-Olson, A.M., et al.: Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. Journal of the American College of Cardiology **75**(7), 722–733 (2020)
13. Kumar, A., Rathor, K., Vaddi, S., Patel, D., Vanjarapu, P., Maddi, M.: Ecg based early heart attack prediction using neural networks. In: 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). pp. 1080–1083. IEEE (2022)
14. Lee, D.C., Markl, M., Dall'Armellina, E., Han, Y., Kozerke, S., Kuehne, T., Nielles-Vallespin, S., Messroghli, D., Patel, A., Schaeffter, T., et al.: The growth and evolution of cardiovascular magnetic resonance: a 20-year history of the society for cardiovascular magnetic resonance (scmr) annual scientific sessions. Journal of Cardiovascular Magnetic Resonance **20**(1), 8 (2018)
15. Liu, C., Wan, Z., Cheng, S., Zhang, M., Arcucci, R.: Etp: Learning transferable ecg representations via ecg-text pre-training. arXiv preprint arXiv:2309.07145 (2023)
16. Liu, X., Wang, H., Li, Z., Qin, L.: Deep learning in ecg diagnosis: A review. Knowledge-Based Systems **227**, 107187 (2021)
17. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
18. Mensah, G.A., Roth, G.A., Fuster, V.: The global burden of cardiovascular diseases and risk factors: 2020 and beyond (2019)
19. Qiu, J., Zhu, J., Liu, S., Han, W., Zhang, J., Duan, C., Rosenberg, M.A., Liu, E., Weber, D., Zhao, D.: Automated cardiovascular record retrieval by multimodal learning between electrocardiogram and clinical report. In: Machine Learning for Health (ML4H). pp. 480–497. PMLR (2023)
20. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
21. Radhakrishnan, A., Friedman, S.F., Khurshid, S., Ng, K., Batra, P., Lubitz, S.A., Philippakis, A.A., Uhler, C.: Cross-modal autoencoder framework learns holistic representations of cardiovascular state. Nature Communications **14**(1), 2436 (2023)
22. Sarkar, P., Etemad, A.: Self-supervised ecg representation learning for emotion recognition. IEEE Transactions on Affective Computing **13**(3), 1541–1554 (2020)
23. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)

24. Siontis, K.C., Liu, K., Bos, J.M., Attia, Z.I., Cohen-Shelly, M., Arruda-Olson, A.M., Farahani, N.Z., Friedman, P.A., Noseworthy, P.A., Ackerman, M.J.: Detection of hypertrophic cardiomyopathy by an artificial intelligence electrocardiogram in children and adolescents. International Journal of Cardiology **340**, 42–47 (2021)
25. Siontis, K.C., Noseworthy, P.A., Attia, Z.I., Friedman, P.A.: Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology **18**(7), 465–478 (2021)
26. Wang, C., Li, Y., Lv, J., Jin, J., Hu, X., Kuang, X., Chen, W., Wang, H.: Recommendation for cardiac magnetic resonance imaging-based phenotypic study: imaging part. Phenomics **1**, 151–170 (2021)
27. Xie, L., Li, Z., Zhou, Y., He, Y., Zhu, J.: Computational diagnostic techniques for electrocardiogram signal analysis. Sensors **20**(21), 6318 (2020)
28. Zhang, H., Liu, W., Shi, J., Chang, S., Wang, H., He, J., Huang, Q.: Maefe: Masked autoencoders family of electrocardiogram for self-supervised pretraining and transfer learning. IEEE Transactions on Instrumentation and Measurement **72**, 1–15 (2022)
29. Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., Langlotz, C.P.: Contrastive learning of medical visual representations from paired images and text. In: Machine Learning for Healthcare Conference. pp. 2–25. PMLR (2022)