A Survey of Augmentation Techniques for Enhancing ECG Representation Through Self-Supervised Contrastive Learning

Deekshith Dade[1], Jake A Bergquist[1,2,3], Rob S MacLeod[1,2,3], Xiangyang Ye[4], Ravi Ranjan[2,3,4], Benjamin A Steinberg[4], Tolga Tasdizen[1,5]

[1] Scientific Computing and Imaging Institute, University of Utah, SLC, UT, USA
[2] Nora Eccles Treadwell CVRTI, University of Utah, SLC, UT, USA
[3] Department of Biomedical Engineering, University of Utah, SLC, UT, USA
[4] School of Medicine, University of Utah, SLC, UT, USA
[5] Department of Electrical and Computer Engineering, University of Utah, SLC, UT, USA

Abstract

The electrocardiogram (ECG) is the most common non-invasive tool to measure the electrical activity of the heart and assess cardiac health. Despite their ubiquity and utility, traditional ECG analysis methods are limited in many impactful diseases. Machine learning tools can be employed to automate task-specific detection of diseases, and to detect patterns that are ignored by traditional ECG analysis. Contemporary machine learning tools are limited by requirements for large labeled datasets, which can be scarce for rare diseases. Self-supervised learning (SSL) can address this data scarcity. We implemented the momentum contrast(MoCo) framework, a form of SSL, using a large clinical ECG dataset. We then assessed the learning using Low Left Ventricular Ejection Fraction (LVEF) Detection as the downstream task. We compared the SSL improvement of LVEF classification across different input augmentations. We observed that optimal augmentation hyperparameters varied substantially based on the training dataset size, indicating that augmentation strategies may need to be tuned based on problem and dataset size.

1. Introduction

The electrocardiogram (ECG) is a foundational tool in diagnosing cardiovascular conditions[1]. With the growth in the volume of ECG data, machine learning techniques present a promising way to boost ECG diagnostic accuracy in various diseases. Innovative applications of machine learning to ECG analysis could extend the utility of ECG beyond its conventional scope, uncovering cardiac irregularities and conditions previously unavailable to traditional analysis.

Machine learning (ML) tools are a powerful solution to automate disease-specific detection tasks with greater precision and efficiency. However, conventional machine learning techniques rely on extensive labeled datasets, limiting their application. The reliability of ML tools depends heavily on the quality and size of the dataset. This is not a problem for common diseases as the data is more readily available. However, the scarcity of labeled data presents a considerable obstacle for rare diseases.

Self-supervised learning (SSL) is a technique that shows promise in alleviating the need for such large labeled datasets. SSL is a process in which the model learns valuable representations/features by comparing data samples without any labels. The learned features/representations of this network are general and can be transferred to a wide range of downstream tasks with substantially less labeled data compared to traditional machine learning methods.

Augmentations play a crucial role in many SSL techniques.[2][3] The goal of these augmentations is to alter the data in a way that enhances the model's ability to recognize and distinguish subtle nuances within signals. We explore various augmentations with a wide range of hyperparameters that the model might encounter in real-world scenarios and measure the performance to learn the scope of these augmentations in ECG representation learning.

Following the augmentation and pretraining phase, the learning was evaluated by focusing on Low LVEF Detection as the downstream task. In this evaluation phase, the performance of the network utilizing the pre-trained weights was evaluated by measuring the Area under the Receiver Operator Characteristic (ROC) Curve (AUC).

2. Methods

Dataset: Digital ECG recordings were collected from 24,868 University of Utah Health patients from 2012 to 2021. Each ECG measured has 8 leads(L1, L2, V1 through

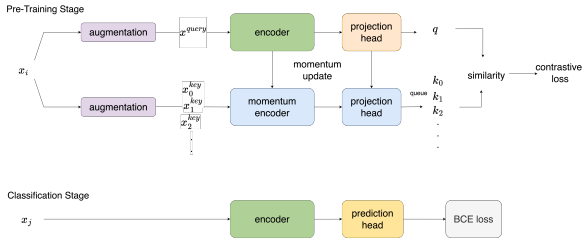

Figure 1. Encoder learns valuable representations during the pretraining stage, and then the weights from the encoder are used to initialize the network to finetune for downstream tasks in the classification stage

V6) and consists of 10 seconds of continuous simultaneous recording from each lead at 500 Hz. Also, each ECG is associated with an ejection fraction measurement within 4 weeks of the ECG recording. For the purpose of training and validating our model, the 24,868 patients were split as 90%(22,381 patients) for pretraining and 1%(223 patients), 5%(1119 patients), 10%(2239 patients) of the 90% is randomly selected for fine-tuning for the downstream task of LVEF Classification. 10%(2,487) of the total patients is reserved as a testing set.

Pretraining: For pretraining, MoCo [4] is the contrastive learning framework used for learning ECG representations. MoCo has shown promising results in learning image representations, which has performed well on downstream tasks such as classification, segmentation, and detection. The contrastive learning in MoCo is framed as a dictionary-lookup task. The dictionary is a dynamic queue of fixed size, and a moving-averaged encoder is used to update the dictionary. The moving-average encoder is updated based on the momentum of the main encoder, which is updated by the backpropagation of gradients. The encoder and the momentum encoder are then challenged to produce representations with low contrastive loss for positive keys and high contrastive loss for all the negative keys in the dictionary. A query and the key are considered positive if they are augmentations of the same image and negative if not.

Augmentations: Our challenge then comes in choosing augmentations. The kind of augmentation and its strength play a crucial role in challenging the encoder to look at contrastive views of the data. Various questions can be raised when choosing augmentations. If the augmentation poses a good enough challenge for the encoder to learn representations that cluster similar signals in latent space and the augmentation's strength, whether it is weak or strong, for the encoder to learn features or to deviate completely from the task at hand.

Few augmentations have been selected for our task where data augmentation has been explored for time series classification in general[2], and the augmentations have been tested specifically for ECG representation learning using contrastive framework[3]. The augmentations that have been chosen to test are:

Gaussian Noise: One of the simple transformations is adding noise to time series data. A random noise of shape of the input signal is picked from $\mathcal{N}(\mu,\ \sigma^{\in})$, where μ is the mean and the standard deviation σ are the hyperparameters being tested.

Gaussian Blur: Gaussian blur is achieved by convolving the image with a Gaussian filter, which is characterized by its standard deviation, σ.

Scaling: Scaling is a multiplication of scaling factor determined by the $\mathcal{N}(\mu,\ \sigma^{\in})$, to the ECG signal lead by lead. Here μ is fixed to 1, and the standard deviation σ is the hyperparameter tested.

Magnitude Warping: The scaling factor is multiplied by the signal so as to warp the signal's magnitude by a smooth curve. The scaling factor is generated by interpolating a cubic spline with knots, where the knots are taken from a Gaussian distribution $\mathcal{N}(\mu,\ \sigma^{\in})$ where μ is fixed to be 1, and the number of Knots, standard deviation σ are the hyperparameters being experimented.

Baseline Warping: The noise to the signal warps the signal's magnitude in a smooth curve. The noise then is the interpolation of a cubic spline with knots, where the knots are taken from a $\mathcal{N}(\mu,\ \sigma^{\in})$ where μ is fixed to be 1 and the number of Knots, standard deviation σ are the hyperparameters.

Time Warping: The warping here happens in the temporal dimension. The augmented time series is of the form:

$$x_i = x_{\tau(1)}, ..., x_{\tau(t)}, ..., x_{\tau(T)},$$

where $\tau(.)$ is a warping function that warps the time steps based on a smooth curve. The smooth curve is defined by a cubic spline with knots. The height of the knots is taken from $\mathcal{N}(\mu, \sigma^{\in})$. Here μ and σ are fixed to 1 and 0.2, and the number of knots is the hyperparameter [2].

Window Warping: Window warping takes a random window of the time series of window ratio and randomly either stretches it by a factor of 2 or contracts it by $\frac{1}{2}$. Here, the window ratio is experimented with.

Encoder: Encoder plays an important role in self-supervised learning by transforming the input data into a lower dimensional feature space. The Spatiotemporal encoder used here is the one used for comparing the performance of LVEF detection using Individual ECG leads [5]. The Spatiotemporal encoder consists of an input stage, temporal and spatial residual blocks, and an output stage. The spatial residual block uses 7×1 convolution filters, and the temporal uses 1×3 filters. The features from the two residual blocks are concatenated before the output stage.

Instead of using a fully connected layer in the output stage as used in the model[5], a projection head that projects the output to 128 dimensions or a prediction head that outputs a 1×1 final output is used, if the task is pretraining or classification, respectively.

Training Details: As mentioned above, 90% (24,868 patients) of all the patients are used for pretraining. For these 24,868 patients, the data contain multiple ECGs for a few patients, resulting in a total of 36,519 ECG signals. For each pretraining task, Loss, Top1-accuracy, and Top-5 accuracy are measured to check the progress of the learning, and for every epoch, the encoder with the best Top1-accuracy is saved. The weights of the model with the best Top1 accuracy are then used as the initialization for the downstream classification task.

Classification: To test the pre-trained model's ability, a binary classification problem of low LVEF detection has been chosen as the downstream task where below 40% LVEF is seen as low [5].

There are two classification stages, finetuning and baseline. For the finetuning stage, the network is initialized with pre-trained weights from self-supervised task and in the baseline stage, the network is initialized randomly. Now, 1% (223 patients), 5% (1119 patients), and 10% (2239 patients) of the training data used for pretraining are used to train the model. The supervised, finetuned model is validated against a test set of 10% (2,487 patients). In the case of classification the ECG that is the one closest in time to the ejection fraction (EF) measurement was used. For all the classification tasks, the "area under the ROC curve" (AUC) is measured, which measures the crucial aspect of the model's ability to discriminate between the positive and negative classes.

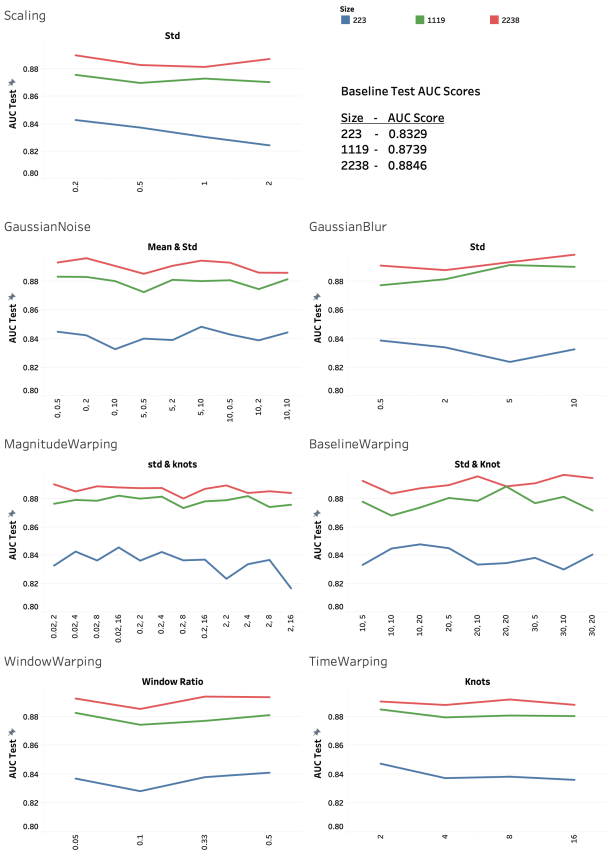

Figure 2. Performance of various augmentations for the downstream task of low LVEF detection. The metric measured here is AUC Score on the validation split

3. Results

The collection of graphs presented in figure 2 presents the performance of various data augmentation techniques on the downstream task of detecting low LVEF from ECG data, with performance measured by the metric AUC. For each augmentation technique, the graphs show the relationship between augmentation intensities and the resulting model performance.

Gaussian noise improves the model's robustness to a point, with medium levels of noise (around 5 mean) performing best, especially for larger datasets. However, performance drops when the noise is too high (around 10 mean). Adding Gaussian blur to the data shows a trend where a moderate amount of blur (standard deviation of around 5) improves the model performance, particularly for large datasets. There is a contrasting trend in the large dataset compared to the small dataset, where the trends are reversed as the noise increases. Scaling augmentation positively impacts the AUC score across all dataset sizes when the standard deviation is low. There is a threshold after which increasing the scaling standard deviation results in a decrease in performance, particularly noticeable in the smaller dataset.

The graph indicates that a moderate amount of standard deviation and knots in magnitude warping benefits the model's performance, with detrimental effects as these increase. Larger datasets show the least sensitivity to changes in parameters. Altering the window ratio has a notable impact on performance, with a clear peak at a window ratio of around 0.5 for all the dataset sizes. Baseline warping shows a consistent trend for larger datasets where the performance peaks at a moderate level of standard deviation of around 20 before declining, whereas the smaller dataset shows an increase in performance at standard deviation 10 before declining. Increasing the number of knots in time warping does not have an impact on AUC scores.

Even though all augmentation techniques do not demonstrate that there is a perfect range in parameter selection

Augmentation	Size	Max	Min	Baseline
Scaling	223	0.8428	0.8243	0.8329
	1119	0.8758	0.8699	0.8739
	2238	0.8901	0.8816	0.8846
GaussianNoise	223	0.8484	0.8327	0.8329
	1119	0.8834	0.8726	0.8739
	2238	0.8963	0.8854	0.8846
GaussianBlur	223	0.8388	0.8238	0.8329
	1119	0.8915	0.8774	0.8739
	2238	0.8988	0.8880	0.8846
MagnitudeWarping	223	0.8456	0.8168	0.8329
	1119	0.8821	0.8734	0.8739
	2238	0.8901	0.8801	0.8846
BaselineWarping	223	0.8478	0.8301	0.8329
	1119	0.8886	0.8681	0.8739
	2238	0.8968	0.8836	0.8846
WindowWarping	223	0.8411	0.8283	0.8329
	1119	0.8825	0.8743	0.8739
	2238	0.8938	0.8852	0.8846
TimeWarping	223	0.8473	0.8361	0.8329
	1119	0.8849	0.8794	0.8739
	2238	0.8918	0.8880	0.8846

Table 1. Comparison of Test AUC Scores Across Different Augmentations and Dataset Sizes with Baseline

that maximizes performance across all dataset sizes, there is an optimal point depending on the size of the dataset. The largest datasets tend to show less sensitivity to augmentation parameters. As presented in Table 1, the results obtained from the implementation of various data augmentation techniques in our study, contrary to literary expectations, did not markedly enhance the model's ability to discern patterns related to low LVEF compared to baseline.

4. Discussion and Conclusions

In this study, we explored self-supervised learning in the context of electrocardiogram (ECG) signal analysis, leveraging the momentum contrast (MoCo) framework. By treating each ECG as an instance, we utilized MoCo's instance discrimination task to learn representations by contrasting positive pairs against a queue of negative samples.

The core of our research deals with a comprehensive survey of augmentation techniques. We identified and evaluated several augmentations, hypothesizing that certain transformations would yield more informative representations by challenging the model to differentiate between clinically relevant signal variations and noise. Through this process, we aimed to discover augmentation strategies that would significantly improve the model's ability to learn generalizable features from ECG signals that can be transferred to various downstream tasks.

By finetuning our pretrained model on a downstream task of detecting low LVEF, we were able to draw conclusions about the utility of the learned representations. Our results demonstrated the nuanced impact of different augmentation techniques on the model's performance. We observed that the dataset size played a pivotal role in determining the effectiveness of each augmentation, with larger datasets showing less sensitivity to augmentations.

Our study has limitations that we would like to explore in future work. An increase in dataset size could lead to learning of a wider array of representations, as seen in MoCo applications with large-scale datasets such as ImageNet and Instagram-1B[4]. The augmentations used in our study were not designed with ECG physiology in mind. A combination of augmentaions was also not explored that could lead to complimentary or deterimental effects. We would like to examine whether these augmentations are equally beneficial when applied to different encoder architectures, which could further validate the robustness of the augmentations for ECG representation learning. We would also like to explore other contrastive learning frameworks. While MoCo provided a solid foundation for our study, but various other frameworks warrant investigation.

Acknowledgments

Support for this research came from the Center for Integrative Biomedical Computing (www.sci.utah.edu/cibc), NIH/NIGMS grants P41 GM103545 and R24 GM136986, NIH/NIBIB grant U24EB029012, and the Nora Eccles Harrison Foundation for Cardiovascular Research.

References

[1] Bergquist JA, Rupp L, Zenger B, Brundage J, Busatto A, MacLeod R. Body surface potential mapping: Contemporary applications and future perspectives. Hearts 2021;2:514–542.

[2] Iwana BK, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. Plos One 2021;16(7):e0254841.

[3] Soltanieh S, Etemad A, Hashemi J. Analysis of augmentations for contrastive ecg representation learning. In 2022 International Joint Conference on Neural Networks (IJCNN). 2022; 1–10.

[4] He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020; 9726–9735.

[5] Bergquist JA, Zenger B, Brundage J, MacLeod RS, Shah R, Ye X, Lyones A, Ranjan R, Tasdizen T, Bunch TJ, et al. Comparison of machine learning detection of low left ventricular ejection fraction using individual ecg leads. In 2023 Computing in Cardiology (CinC), volume 50. IEEE, 2023; 1–4.

Address for correspondence:

Deekshith Dade
University of Utah
72 Central Campus Dr, Salt Lake City, UT 84112
deekshith.dade@utah.edu