Impact of Pre-Processing Decisions on Automated ECG Classification Accuracy

Adrian K Cornely, Grace M Mirsky

Benedictine University, Lisle, IL, USA

Abstract

Electrocardiography is well established as an effective clinical tool for detection and diagnosis of cardiac arrhythmias and abnormalities. The objective of the 2021 PhysioNet/Computing in Cardiology Challenge was for teams to develop automated classification algorithms for reduced-lead ECGs. While it is well-known that proper pre-processing is very important for the success of classification algorithms, there is not universal agreement as to the appropriate pre-processing steps for automated ECG classification. Papers from the top 15 finishers in the Challenge as well as the bottom ten finishers were examined to determine what pre-processing steps were applied by each team.

The most commonly used pre-processing steps included resampling to a consistent sampling rate, applying a bandpass filter, normalizing and using a fixed signal length. There were a number of similarities in the preprocessing steps used by the top 15 teams, whereas all of these steps were not applied in the majority of approaches for the bottom ten teams. In the bottom ten participants, less than half used a bandpass filter, and only three applied some type of normalization. This investigation underscores the importance of appropriate pre-processing for strong classification accuracy and the need for a universal approach to pre-processing techniques in automated ECG classification.

1. Introduction

Electrocardiography is well-established as an effective clinical tool for detection and diagnosis of cardiac arrhythmias and abnormalities [1]. Due to the equipment cost and required staff expertise, however, not all medical facilities worldwide are capable of regularly providing patients with standard 12-lead electrocardiograms (ECGs). It is currently unclear if fewer leads could potentially provide comparable classification accuracy as the standard 12-lead ECG. As a result, the objective of the 2021 PhysioNet/Computing in Cardiology Challenge was for teams to develop automated classification algorithms for reduced-lead ECGs on a large dataset acquired from several geographically separated sites [2,3].

Over 200 teams attempted the Challenge, utilizing various algorithms for classification, including neural networks [4], wavelet transforms [5] and LSTM networks [6]. Transfer learning was also used by some teams [7,8]. Prior to the Challenge, various degrees of success have been achieved in the reconstruction accuracy of 12-lead ECGs from reduced lead sets [9]. While some studies have shown that patient-specific lead derivation provided better accuracy than generalized derivations [10] (except in the presence of ischemic events [11]), Independent Component Analysis (ICA) has been shown to accurately reconstruct precordial leads [12]. Additional studies have not reproduced comparable accuracy, however, in the presence of rhythm and morphological abnormalities [13]. Various prior studies, which have been limited in terms of dataset availability and reproducibility, have motivated the 2021 Challenge in order to rigorously determine the potential classification accuracy achievable for various reduced lead set combinations.

In addition to the many papers resulting from the Challenge itself, substantial focus in the recent literature has been placed on the accuracy and potential clinical utility of different machine learning models for automated ECG classification. However, while it is well-known that proper pre-processing is very important for the success of classification algorithms, there is not universal agreement as to the appropriate pre-processing steps for automated ECG classification. The structure of the Challenge provides the opportunity to better investigate and compare the various options for pre-processing for different algorithms applied to the same large-scale training dataset.

2. Methods

We reviewed the Computing in Cardiology Conference papers from the top fifteen finishers in the Challenge to determine what pre-processing steps were applied by each team. We first looked at resampling, which allows signals to be compared at the same rate. We next examined normalization techniques, which are useful in order to

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.252

compare signals on the same scale in terms of magnitude. Filtering/noise reduction was also observed, which is typically done to reduce noise while maintaining physiologically meaningful frequencies in the signals. Next, we looked at fixed signal lengths, which are necessary for certain machine learning algorithms. Finally, Challenge data that are not used for training or external datasets which are added to the training data are examined. Datasets which have atypical characteristics (very long signals, low sampling rate, etc.) or with ambiguous classification labels may be discarded, whereas additional data may be added to supplement and diversify the training set.

In order to assess the generalizability of the algorithms, we examined the standard deviation of the scores on the different test sets: China Physiological Signal Challenge (CPSC), Georgia (G12EC), Undisclosed and the University of Michigan (UMich). Challenge participants were given training data from two of the datasets, CPSC and G12EC, but these data were non-overlapping with the test and validation data. However, the other two datasets, Undisclosed and UMich, only appeared in the test set, so participants were unable to train with records from either of these sites. As a result, the generalizability can be ascertained from the standard deviation by observing the consistency of the participant submissions across these different test sets, which may be more meaningful in terms of practical clinical utility of these algorithms rather than overall performance.

3. Results

There were a number of similarities in the pre-processing steps used by the top 15 teams. All challengers in this group used a consistent sampling rate across all input data. Seven teams used 500Hz, 3 teams used 250Hz, and the remaining teams used 300Hz, 257Hz, 200/100Hz, 150Hz, and 125Hz, respectively. Over half used some type of normalization or standardization of the signals, including min-max and z-score normalization. Seven challengers applied bandpass filters, with varying passband ranges. All but two of the fifteen entries used a fixed signal length, though the actual length varied, with the smallest being 2000 samples and the largest being 15000 samples. Five teams removed the PTB and INCART datasets from the training set. Two teams removed signals based on length; one team removed signals longer than 15 seconds, and another team removed signals longer than 20 seconds.

Interestingly, the vast majority of the top fifteen participants used neural networks with residuals, whereas only one of the bottom ten participant teams used a version of this algorithm. In addition, in the bottom ten participants, less than half used a bandpass filter, and only three applied some type of normalization. Six of the participants in the bottom ten group used a fixed sampling rate, and all but one of the teams used a fixed signal length, which was typically achieved by cropping or zero-padding the signals.

The official ranking order did not fully correlate with the relative performance between submissions on the hidden test subsets, as shown in Figure 1 and Table 1. Some participants did worse against the blind set relative to other algorithms which ranked lower. For instance, one of the 5th place teams scored 0.89 on CPSC but received a score of 0.31 on the undisclosed set. On the other hand, the other team that tied for 5th place scored 0.59 on CPSC and received a score of 0.42 on the undisclosed set.

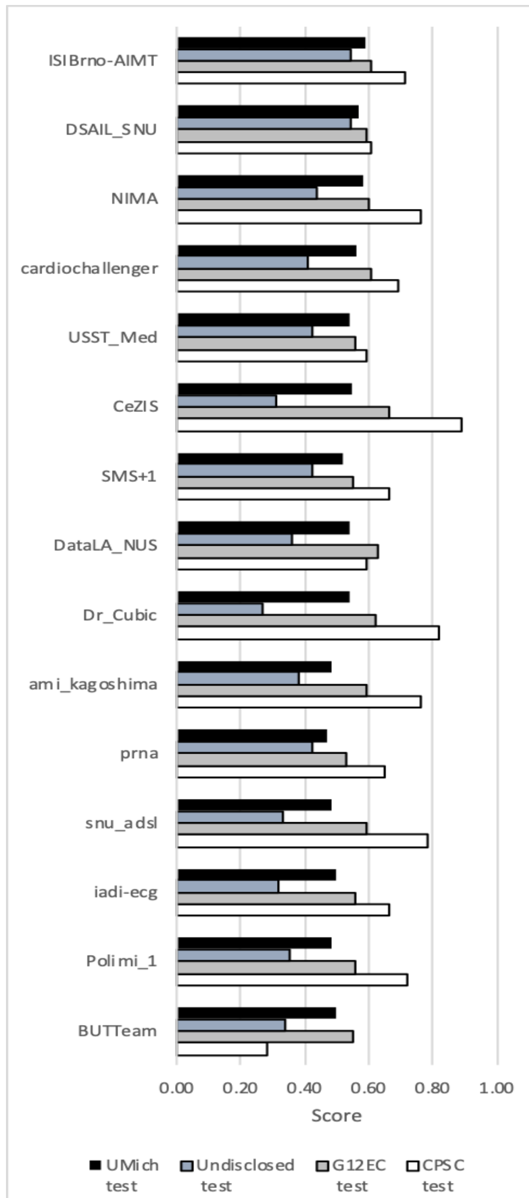

Figure 1. Official scores for individual test sets per team.

Table 1. Standard deviation across the four test sets.

Team Name	Ranking	Standard Deviation
ISIBrno-AIMT	1	0.0714
DSAIL_SNU	2	0.0299
NIMA	3	0.1310
cardiochallenger	4	0.1179
USST_Med	5	0.0746
CeZIS	5	0.2410
SMS+1	7	0.0988
DataLA_NUS	7	0.1192
Dr_Cubic	9	0.2278
ami_kagoshima	10	0.1628
prna	11	0.0991
snu_adsl	11	0.1895
iadi-ecg	11	0.1428
Polimi_1	11	0.1548
BUTTeam	15	0.1282

4. Discussion and Conclusions

While these results predictably reinforce the well-known need to properly pre-process data for machine learning algorithms, they also highlight the importance of a generally agreed upon standard for critical pre-processing steps. A standard set of steps would be particularly beneficial to advance this important area of research, and it is clear that such an agreement has not yet been reached due to the significant variability in pre-processing approaches among the Challenge submissions, which have resulted in varying degrees of classification accuracy among Challenge participants.

From observing the pre-processing steps taken in the top fifteen finishers compared with the bottom ten finishers, some obvious patterns emerge. In particular, it appears that the combination of residual neural networks with a fixed sampling rate and fixed signal length is very effective at achieving high classification accuracy. However, there is an interesting case [28] in the bottom ten in which ResNet, a fixed sampling rate and a fixed signal length are used, but the classification accuracy is significantly worse than other participants with similar approaches. We speculate that although a fixed sampling rate was used, the sampling rate chosen is relatively low, 64Hz, which may have contributed to the reduction in classification accuracy. In particular, that sampling rate is too low to fully capture the generally accepted frequency range of a standard electrocardiogram, i.e. 0.05-150Hz, which implies that the sampling rate used by this team may have omitted relevant signal components at higher frequencies.

Based on the standard deviation of algorithmic performance on the test sets, it is clear that the official Challenge rankings are not necessarily indicative of the generalizability of the individual approaches. For example, for team CeZIS, which was tied for fifth place in the official rankings, the model was likely overtrained on the known datasets (CPSC and G12C), achieving the highest score of any algorithm in the top 15 on the CPSC dataset. Even though the overall Challenge score and ranking were among the top performers overall for this team, this algorithm performed much worse on previously unseen datasets than other comparably ranked algorithms, resulting in a relatively high standard deviation across all four test sets. Therefore, the reduced classification accuracy on the previously unseen data speaks to the lack of robust generalizability of the algorithm. In comparison, the other team tied for fifth place, USST_Med, had a much smaller overall standard deviation across the four datasets, so even though their ranking was the same as CeZIS, this analysis implies that the USST_Med algorithm is much more likely to accurately classify previously unseen data since it does not appear to be overtrained on the provided training sets.

On the other hand, the second place team in the Challenge, DSAIL_SNU, had the lowest overall standard deviation. It also tied for the highest score on the undisclosed set, with a value of 0.54. However, it did not have the highest classification accuracy for any other individual dataset, even with its very high overall score. These factors indicate that this algorithm is generalizable and highly likely to accurately classify previously unseen datasets.

There are a number of limitations of this analysis, most notably the significant variations in overall approach between the participants, which make it impossible to fully isolate the effects of the individual pre-processing steps on classification accuracy.

References

[1] Kligfield P et al. "Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: the Electrocardiogram and Its Technology a Scientific Statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology," *Journal of the American College of Cardiology*, vol. 49, no. 10, pp. 1109-1127, 2007.

[2] Reyna MA et al. "Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021". *Computing in Cardiology*, 48: 1-4, 2021.

[3] Alday EAP et al., "Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Measurement*, 41, no. 12 (2020): 124003. doi: 10.1088/1361-6579/abc960

[4] Nejedly P et al. “Classification of ECG using Ensemble of Residual CNNs with Attention Mechanism,” *Computing in Cardiology*, 48: 1-4, 2021.
[5] Natarajan A et al. “Convolution-Free Waveform Transformers for Multi-Lead ECG Classification,” *Computing in Cardiology*, 48: 1-4, 2021.
[6] Pan L et al. “MTFNet: A Morphological and Temporal Features Network for Multiple Leads ECG Classification,” *Computing in Cardiology*, 48: 1-4, 2021.
[7] Bortolan G. “3-D ECG images with Deep Learning Approach for Identification of Cardiac Abnormalities from a Variable Number of Leads,” *Computing in Cardiology*, 48: 1-4, 2021.
[8] Cornely AK et al. “Reduced-Lead Electrocardiogram Classification using Wavelet Analysis and Deep Learning,” *Computing in Cardiology*, 48: 1-4, 2021.
[9] Tomašić I, Trobec R. "Electrocardiographic Systems with Reduced Numbers of Leads—Synthesis of the 12-lead ECG." *IEEE reviews in biomedical engineering*, 7: 126-142, 2013.
[10] Nelwan SP et al. "Minimal Lead Sets for Reconstruction of 12-lead Electrocardiograms." *Journal of electrocardiology*, 33: 163-166, 2000.
[11] Guldenring D et al. "Estimation Accuracy of a Reduced Lead System During Simulated Ischemia," *Computing in Cardiology*, 38: 237-240, 2011.
[12] Ostertag, MH, Tsouri, GR. "Reconstructing ECG Precordial Leads from a Reduced Lead Set Using Independent Component Analysis." *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011.
[13] Gregg RE et al. “Where Do Derived Precordial Leads Fail?” *J Electrocardiol*, 41(6):546-52, 2008.
[14] Han H et al. “Towards High Generalization Performance on Electrocardiogram Classification,” *Computing in Cardiology*, 48: 1-4, 2021.
[15] Wickramasinghe NL, Athif M. “Multi-label Cardiac Abnormality Classification from Electrocardiogram using Deep Convolutional Neural Networks,” *Computing in Cardiology*, 48: 1-4, 2021.
[16] Srivastava A et al. “Channel Self-Attention Deep Learning Framework for Multi-Cardiac Abnormality Diagnosis from Varied-lead ECG Signals,” *Computing in Cardiology*, 48: 1-4, 2021.
[17] Antoni L et al. “A Two-Phase Multilabel ECG Classification Using One-Dimensional Convolutional Neural Network and Modified Labels,” *Computing in Cardiology*, 48: 1-4, 2021.
[18] Cai W et al. “Classifying Different Dimensional ECGs Using Deep Residual Convolutional Neural Networks,” *Computing in Cardiology*, 48: 1-4, 2021.
[19] Ren H et al. “Robust and Task-Aware Training of Deep Residual Networks for Varying-Lead ECG Classification,” *Computing in Cardiology*, 48: 1-4, 2021.
[20] Vazquez CG et al. “Two Will Do: Convolutional Neural Network with Asymmetric Loss, Self-Learning Label Correction, and Hand-Crafted Features for Imbalanced Multi-Label ECG Data Classification,” *Computing in Cardiology*, 48: 1-4, 2021.
[21] Li X et al. “Towards Generalization of Cardiac Abnormality Classification Using ECG Signal,” *Computing in Cardiology*, 48: 1-4, 2021.
[22] Seki H et al. “Reduced-Lead ECG Classifier Model Trained with DivideMix and Model Ensemble,” *Computing in Cardiology*, 48: 1-4, 2021.
[23] Aublin P et al. “Cardiac Abnormality Detection based on an Ensemble Voting of Single-Lead Classifier Predictions,” *Computing in Cardiology*, 48: 1-4, 2021.
[24] Muscato FM et al. “Ensemble Learning of Modified Residual Networks for Classifying ECG with Different Set of Leads,” *Computing in Cardiology*, 48: 1-4, 2021.
[25] Suh J et al. “Learning ECG Representations for Multi-Label Classification of Cardiac Abnormalities,” *Computing in Cardiology*, 48: 1-4, 2021.
[26] Vicar T et al. “Cardiac Abnormalities Recognition in ECG Using a Convolutional Network with Attention and Input with an Adaptable Number of Leads,” *Computing in Cardiology*, 48: 1-4, 2021.
[27] Linschmann O et al. “Multi-label Classification of Cardiac Abnormalities for Multi-lead ECG Recordings Based on Auto-encoder Features and a Neural Network Classifier,” *Computing in Cardiology*, 48: 1-4, 2021.
[28] Alkhodari M et al. “Swarm Decomposition Enhances the Discrimination of Cardiac Arrhythmias in Varied-lead ECG Using ResNet-BiLSTM Network Activations,” *Computing in Cardiology*, 48: 1-4, 2021.
[29] Krivenko SS et al. “Using Mel-frequency Cepstrum and Amplitude-Time Heart Variability as XGBoost Handcrafted Features for Heart Disease Detection,” *Computing in Cardiology*, 48: 1-4, 2021.
[30] Uhlemann T et al. “An Ensemble Learning Approach to Detect Cardiac Abnormalities in ECG Data Irrespective of Lead Availability,” *Computing in Cardiology*, 48: 1-4, 2021.
[31] Warrick PA et al. “Arrhythmia Classification of Reduced-Lead Electrocardiograms by Scattering-Recurrent Networks,” *Computing in Cardiology*, 48: 1-4, 2021.
[32] Ignacio PS. “Leveraging Period-specific Variations in ECG Topology for Classification Tasks,” *Computing in Cardiology*, 48: 1-4, 2021.
[33] Osnabrugge N et al. “Multi-label Classification on 12, 6, 4, 3 and 2 Lead Electrocardiography Signals using Convolutional Recurrent Neural Networks,” *Computing in Cardiology*, 48: 1-4, 2021.
[34] Baumgartner M et al. “ECG Classification Combining Conventional Signal Analysis, Random Forests and Neural Networks – a Stacked Learning Scheme,” *Computing in Cardiology*, 48: 1-4, 2021.
[35] Singstad B-J et al. “Multi-label ECG Classification using Convolutional Neural Networks in a Classifier Chain,” *Computing in Cardiology*, 48: 1-4, 2021.
[36] Puszkarski B et al. “N-BEATS for Heart Dysfunction Classification,” *Computing in Cardiology*, 48: 1-4, 2021.

Address for correspondence:

Grace M. Mirsky
Benedictine University
5700 College Road
Lisle, IL 60532, USA
gmirsky@ben.edu