

心アミロイドーシス早期発見を目的とした心電図データ拡張手法の検討 Exploring ECG Data Augmentation Methods for Early Detection of Cardiac Amyloidosis

椎名 隆太^{*1} 大和 淳司^{*1} 三上 弾^{*1} 山本 博之^{*2}

Ryuta Shiina. Junji Yamato. Dan Mikami. Hiroyuki Yamamoto

^{*1} 工学院大学

^{*2} 東京医科大学病院

Kogakuin University of Technology & Engineering Tokyo Medical University Hospital

Cardiac amyloidosis is a rare disease often diagnosed late due to mild early symptoms. Early detection is crucial, but limited ECG data poses a challenge. This study developed a data expansion method to aid early diagnosis. R-wave peaks were detected, and heartbeats were extracted and parameterized. Differences in R-wave intervals, P- and T-wave timing, and amplitudes between normal and abnormal waveforms were analyzed. New waveforms were generated using linear and spline interpolation. Using the expanded data, models were built with CNN, random forest, and SVM, and their accuracy was evaluated. Results suggest that even with limited data, these methods can support early diagnosis. This study marks a significant step toward addressing data scarcity in rare diseases and developing diagnostic tools. Further validation and clinical application of these methods are planned.

1. 背景

心アミロイドーシス(CA)は、異常なアミロイドタンパクが心筋に蓄積することで心機能に影響を及ぼす疾患であり、早期発見が治療方針の決定や予後改善において極めて重要である。しかし、CAの初期段階では心電図上の特徴が微細であるため、従来の診断法では見逃される可能性がある。

近年、機械学習を活用した診断支援技術が発展し、心電図データを用いた疾患の自動検出が注目されている[Acharya 17, Martis 13, Li 16, Sharma 15, Fujita 17, Safdarian 14, Kojuri 15]。機械学習による分類タスクは数多く提案されているが、これらは一般的に十分なデータ量を前提としている。しかし、心電図データの収集には限界があり、特に希少疾患であるCAにおいては、学習に十分なデータ量を確保することが困難である。このようなデータ不足は、モデルの汎化性能を低下させ、診断精度の向上を妨げる要因となる。

データが少ない疾患に対するアプローチとして、データ拡張が有効であると考えられる。実際に、データ拡張によって分類モデルの精度が向上した例も報告されている[Nicolas 2018]。本研究では、CA患者群およびコントロール群の心電図データを対象に、PQRST波形の特徴を考慮したデータ拡張手法を検討する。特に、R波を基準とした各心拍の時刻情報および振幅情報を用いた拡張手法を提案し、モデルの診断精度向上に寄与するかを評価することを目的とする。

2. 方法

本研究では、12誘導で構成されるCA群およびコントロール群の心電図データを用いて、データ拡張手法の有効性を検証した。以下に、詳細を示す。

2.1 データセット

データセットとして、東京医科大学病院から収集した心電図データを使用した。心電図データはCA群34名およびコントロール群71名、合計105名の患者から得られたものである。また、心電図データはサンプリングレート500Hzの12誘導心電図によって構成されており、各患者の10秒間の心電図データが収

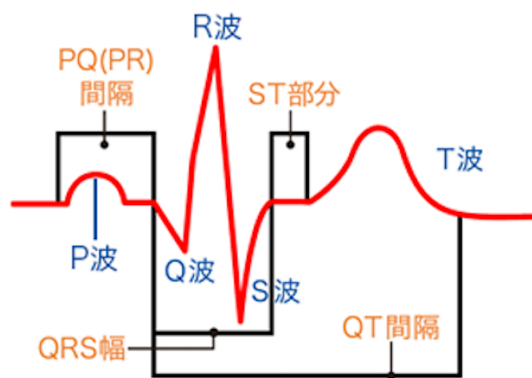


図1: 一般的な心電図の所見

められている。これらの心電図データに含まれるノイズを低減するために、ローパスフィルタによるフィルタリングを行った。

2.2 データ拡張手法

フィルタリング済みの12誘導心電図データを用いて、データの伸縮によるデータ拡張を行った。以下にその手順を示す。また、一般的な心電図の所見を図1に示し、(2), (3)による拡張を実施した波形を図2, 図3, 図4に示す。

(1) 心拍検出とパラメータ算出

12誘導心電図の第2誘導を用いて、10秒間の心電図データにあるRピークを検出した。次に、各Rピークを基準として、P, Q, S, T波の時刻および振幅情報を算出した。これらのパラメータをもとに、P波からT波までの区間を1心拍として定義した。

(2) 振幅(Y軸)の伸縮による拡張

各心拍に対して、振幅を ± 1.1 倍の範囲でスケールリングを行った。振幅のスケールリングにはスプライン補間を用いた。

(3) 時間(X軸)の伸縮による拡張

1心拍内のQ-S区間およびQ-T区間に着目し、各心拍の時間軸を ± 1.1 倍の範囲でスケールリングを行った。時間軸のスケールリングには線形補間を用いた。

(4) 一貫性の維持

(2), (3)のデータ拡張は CA 群およびコントロール群の全ての患者データに対して適用した. この際, 10 秒間の心電図における各心拍は, 全て同一のスケールで拡張されるように処理を行った. これにより, 心拍間で一貫性を保てるようにした. また, 12 誘導全てに対して同一のスケールで拡張を行い, 異なる誘導間のデータの整合性が保たれるようにした.

2.3 精度評価

本研究では, データ拡張手法の有効性を検証するために, CA 群およびコントロール群の 12 誘導心電図データを用いた精度評価を行った. 精度評価には CNN, SVM, ランダムフォレストの 3 つの異なるモデルを使用し, データ拡張の有無が精度向上に寄与するかを検証した. 以下に各モデルの詳細を示す.

CNN(畳み込みニューラルネットワーク)

CNN は画像認識や画像処理に使用される深層学習のモデルの一種である. 本モデルでは, Kachueeらが提案した CNN モデル構造を使用し, 入力層のみを 12 誘導心電図に対応させた [Kachuee 18]. クロスバリデーションには KFold(20 分割)を採用し, 各患者のデータが重複しないようにグループ分けを行った.

SVM(サポートベクターマシン)

SVM は, 高次元データに対しても適用可能であり, マージン最大化による識別を行う分類モデルである. 本モデルでは, 非線形な分類問題に適している RBF カーネルを使用した. また, データの偏りを防ぐため, GroupKFold(20 分割)によるクロスバリデーションを適用し, 同じ患者のデータが学習データとテストデータに重複しないように分割した.

ランダムフォレスト

ランダムフォレストは, 複数の決定木を組み合わせたアンサンブル学習モデルである. 本モデルでは, クラス不均衡を考慮して, クラス間の重みを自動で付けるように設定した. CNN と同様に, クロスバリデーションには KFold(20 分割)を採用し, 各患者のデータが重複しないようにグループ分けを行った.

3. 結果

105 件のデータ(CA 群:34 件, コントロール群:71 件)のうち, パラメータ化が可能だったのはそれぞれ 10 件, 48 件であった. パラメータ化できた波形をもとに, CA 群は 15 倍, コントロール群は 2 倍のデータ拡張を実施し, 元のデータと拡張後のデータを合わせて, 総計 351 件(CA 群:184 件, コントロール群:167 件)のデータを使用した.

3.1 元データと拡張データを用いた精度比較

拡張データがモデルの判別精度にどのように寄与するかを検証するために, 元データのみを使用したモデルと, 拡張データを含むモデルで実験を行った. これらの精度結果を表1, 表 2 に示す. 表1, 表2より, CNN は拡張データを用いたことで Precision, Recall は向上したものの, Accuracy が 4 ポイント低下した. ランダムフォレストについても, Accuracy, Precision は拡張データを使用したことでそれぞれ 11 ポイント, 7 ポイント低下した. 一方, SVM では拡張データの使用により, 精度が全体的に大きく改善しており, 拡張データがモデルの精度向上に寄与することが示唆された.

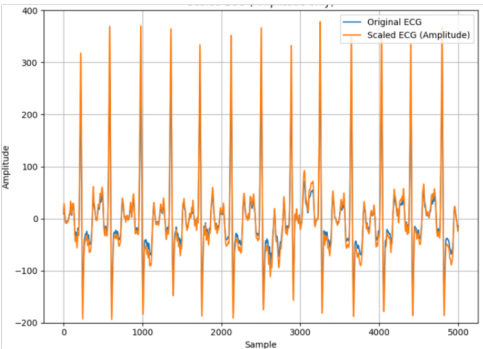


図 2:Y 軸方向の伸縮を実施した波形

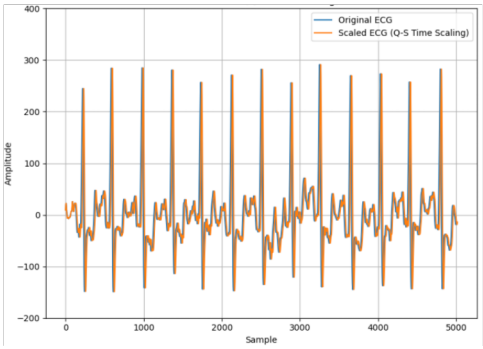


図 3:Q-S 区間の伸縮を実施した波形

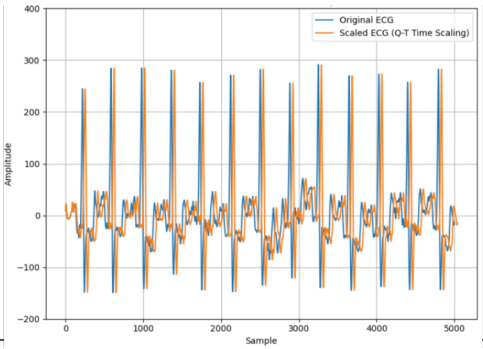


図 4:Q-T 区間の伸縮を実施した波形

表 1:分類精度(元データのみ)

	Accuracy (%)	Precision (%)	Recall (%)
CNN	71	68	60
SVM	58	51	51
RF	70	71	56

表 2:分類精度(元データ+拡張データ)

	Accuracy (%)	Precision (%)	Recall (%)
CNN	67	70	68
SVM	79	83	80
RF	59	64	56

3.2 モデル間の精度比較

本分類タスクにおいて、CNN、SVM、ランダムフォレストのどのモデルが最も適しているかを比較した。表 1、表 2 より、CNN は元データでは最も高い Accuracy を記録したが、拡張データでは他のモデルに比べて Accuracy が低下した。しかし、Recall が向上しており、CA 群の識別能力が向上している可能性がある。次に、ランダムフォレストは元データの方が精度は高いが、拡張データを使うことで精度が低下した。これより、ランダムフォレストには拡張データが適していない可能性があることが示唆された。最後に、SVM は拡張データで大きく改善し、最も高い Accuracy、Precision、Recall を達成した。これは本研究で提案した拡張手法が SVM にとって効果的だったと言える。

4. 考察

4.1 拡張データの有効性

本研究では、3つのモデルによる分類タスクを行い、拡張データの有効性を検証した。その結果、SVM では精度向上に大きく寄与したが、その他のモデルでは性能が低下するという結果であった。この原因として、拡張対象のデータの種類の少ないことが関与していると考えられる。本研究でパラメータ化ができた心電図波形はどれも R ピークが高く、検出が容易であった。しかし、パラメータ化ができなかった心電図波形は、P、Q、R、S、T 波の各ピークの大小関係が、同じ波形の心拍間でも異なる場合があった。そのため、これらの複雑な波形がパラメータ化できず、拡張データの種類の確保できなかった。これより、CA の分類精度を向上させるためには、パラメータ化できなかった複雑な波形も扱う必要があると考えられる。その他にも、拡張時のスケールリングは 0.9-1.1 倍であったが、範囲を調整することで、精度向上を図ることができると考えられる。また、専門医と協力の元、拡張データに適切なラベル付けをすることで、拡張データの有用性が高まると考えられる。

4.2 分類モデルの選択

本研究で提案したデータ拡張手法は、振幅スケールリングと時間スケールリングを用いたものであった。これには SVM のような線形分類器には有効であることが示された。一方で、最も性能が低かったランダムフォレストでは、特徴選択やハイパーパラメータ調整が必要になると考えられる。また、拡張後のデータを用いた特徴選択や次元削減を行うことで、分類精度を改善させることができると考えられる。CNN は Accuracy が低下したものの、Recall は向上しているため、より豊富な拡張データを使用することで、他のモデルよりも性能が向上する可能性があると考えられる。今後は、より多くの拡張データを用意し、CA の早期発見のためのモデル開発に取り組んでいく。

5. まとめ

本研究で提案したデータ拡張手法は、SVM の分類性能を大きく向上させ、CA 群の分類が改善された。しかし、CNN、ランダムフォレストでは性能が一部低下したため、モデルの適用方法およびデータ拡張の最適化について、さらなる検討が必要であることが示唆された。

今後の課題として、パラメータ化できていない複雑な心電図波形を対象にデータ拡張を進めるとともに、スケールリングを調整することが挙げられる。

参考文献

- [Acharya 17] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan, “A deep convolutional neural network model to classify heartbeats,” *Computers in biology and medicine*, vol. 89, pp. 389–396, 2017.
- [Martis 13] R. J. Martis, U. R. Acharya, C. M. Lim, K. Mandana, A. K. Ray, and C. Chakraborty, “Application of higher order cumulant features for cardiac health diagnosis using ecg signals,” *International journal of neural systems*, vol. 23, no. 04, p. 1350014, 2013.
- [Li 16] T. Li and M. Zhou, “Ecg classification using wavelet packet entropy and random forests,” *Entropy*, vol. 18, no. 8, p. 285, 2016.
- [Sharma 15] L. Sharma, R. Tripathy, and S. Dandapat, “Multiscale energy and eigenspace approach to detection and localization of myocardial infarction,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1827–1837, 2015.
- [Fujita 17] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, “Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals,” *Information Sciences*, vol. 415, pp. 190–198, 2017.
- [Safdarian 14] N. Safdarian, N. J. Dabanloo, and G. Attarodi, “A new pattern recognition method for detection and localization of myocardial infarction using t-wave integral and total integral as extracted features from one cycle of ecg signal,” *Journal of Biomedical Science and Engineering*, vol. 7, no. 10, p. 818, 2014.
- [Kojuri 15] J. Kojuri, R. Boostani, P. Dehghani, F. Nowroozpour, and N. Saki, “Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram,” *Journal of Cardiovascular Disease Research Vol*, vol. 6, no. 2, p. 51, 2015.
- [Nicolas 2018] Nicolas Mine, “Model from ArXiv 1805.00794” <https://www.kaggle.com/code/coni57/model-from-arxiv-1805-00794> (参照:2025-2-9)
- [Kachuee 18] Mohammad Kachuee, Shayan Fazeli, Majid Sarrafzadeh, “ECG Heartbeat Classification: A Deep Transferable Representation,” *IEEE, ICHI*, 2018.