

新生児心電図のクラスタリング及びその解釈 Clustering and Interpretation of Neonatal Electrocardiograms

○林夢海¹, 関弘翔², 宮野咲紀³, 佐藤洋明⁴, 細野裕行²

*Yumi Hayashi¹, Hiroto Seki², Saki Miyano³, Hiroaki Sato⁴, Hiroyuki Hosono²

Abstract: In this study, we cluster ECG data using unsupervised learning to identify trends in the maturation process of the autonomic nervous system and causal relationships that could not be classified in humans. In this paper, we examine whether there are individual differences in ECGs by clustering neonatal ECGs using the k-means++ method and interpreting the clustering results using XAI.

1. まえがき

新生児の自律神経系の成熟過程は身体の高さや歩行、会話など目に見えて劇的な変化を実感できる身体成長とは違い、表面上では実感することが困難である。しかし、成人の自律神経系に比べて明らかに未熟であり、身体成長に伴い自律神経系も成熟する。

新生児の自律神経系の成熟過程を把握することで、自律神経系の未熟さが一つの原因である乳幼児突然死症候群(SIDS)などの病気の早期発見につながると考えられる。一般的に、自律神経系の検査には、薬理遮断や採血、針電極による測定といった侵襲的な手法や体位の変換など被験者の積極的な協力を必要とする手法が用いられている。一方、新生児に対しこれらの手法を用いることは比較的困難であるため、より簡単で被験者の負担の少ない非侵襲的な測定方法が望ましい。

先行研究^[1]では、心拍変動から自律神経活動が読み取り可能であることを利用して、新生児の心電図データから自律神経系の成熟過程の把握を検討した。自律神経遮断後、心拍変動がほとんど消失することから自律神経系の影響で心臓の拍動間隔が微妙に変動することが知られている^[2]。この変動には周波数依存性があるため、パターンを捉えることによる自律神経の活動の定量的な評価が機械学習を用いずに試まれたが、共通した発達パターンが認められなかったことが報告されている。

そこで本研究では、心電図データをグラフ化した画像に対する機械学習や深層学習を用いてクラスタリングやパターンの抽出を行い、その結果に基づいた評価を検討する。これにより、自律神経系の成熟過程の傾向や人間では分類できなかった因果関係の洗い出しを目指す。

本稿では、個人によって心電図に違いがあるのかを確認するために2人の新生児心電図データを混在させ、教師なし機械学習によるクラスタリングを検証し、

XAIのSHAP^[3]を用いてクラスタリング結果の理由付けを行った。

2. データクリーニング

新生児のHR(心拍数)の心電図データに対して、はじめにデータクリーニングを実行する。データクリーニングはデータ域抽出、誤検出値の修正、外れ値除去、クラスタリング用画像生成の4段階で行う。Fig. 1にデータクリーニング前後の画像例を示す。この時、データと共にewm_mean(指数移動平均)をプロットした。

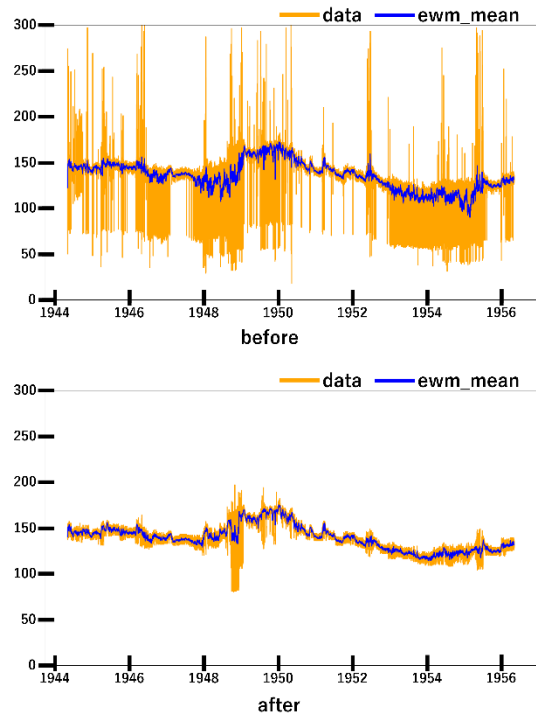


Figure 1. before and after data cleaning

2.1. データ域抽出

欠損値を含み1人当たり500時間以上ある心電図データから12時間の連続値を抽出する。これは3時間毎に食事をとることやサーカディアンリズム等から12時間区切りと設定した。

1: 日大理工・学部・情報, 2: 日大理工・教員・情報, 3: 日大理工・院(後)・情報, 4: 自治医大附属さいたま医療センターNICU

2.2. 誤検出値の修正

外れ値ではなく心拍の誤検出によって本来の値の整数倍で検出された値を修正する。新生児の HR の平均値が 120 であることから、80 以下の値を 2 倍に、200 以上の値を 1/2 倍に修正する。

2.3. 外れ値除去

データを 60 分割してから平均値 μ と標準偏差 σ を求め、 $\mu \pm 2\sigma$ の範囲外の値を外れ値とみなして除去をし、データを再結合する。

2.4. クラスタリング用画像生成

クラスタリング用画像として指数移動平均のみをプロットしたグラフを生成する。指数移動平均は、直近のデータを重視し、データが古くなるほど指数関数的に重みが減少する移動平均の一種である。

3. k-means++法によるクラスタリング

データクリーニングを 2 人の新生児心電図データそれぞれに実行し、2 データを混在させた画像群に対して k を 2 とした k-means++クラスタリングを実行する。Table 1 に、クラスタリング結果を示す。表中の HrsOfAge は生後時間を示している。Table 1 より、A の生後 2115-2127 時間を除いて個人の識別ができたことがわかる。

Table 1. Clustering Results

ID	HrsOfAge	Group number
A	1197-1209	0
	1216-1228	0
	1254-1266	0
	1324-1336	0
	1867-1879	0
	1880-1891	0
	1920-1932	0
	1944-1956	0
B	2115-2127	1
	295-307	1
	319-331	1
	386-398	1
	413-425	1
	484-496	1
	535-547	1
	552-564	1
	564-576	1
	579-591	1

4. SHAP によるクラスタリング結果の解釈

shap モジュール^[4]を用いてクラスタリングにおける画像部位の影響度合いを可視化する。SHAP は、機械学習モデルの出力の根拠を入力への貢献値 (SHAP value) で説明する XAI である。赤いピクセルはクラスの確率を上げる正の SHAP value を表し、青いピクセルはクラスの確率を下げる負の SHAP value を表す。

Fig.2 に SHAP を用いて Table 1 のクラスタリング結果の理由付けを行った結果を示す。これより、group 1

では特にグラフで著しい起伏のある山や谷の頂点に赤いピクセルが分布している。一方、group 0 の結果は、中央付近の波形がない部分に赤いピクセルが分布している。つまり、山・谷がはっきりしているグラフは group 1 の確率が高くなり、あまりはっきりしていないグラフは group 0 に分類される傾向にあると考察できる。



Figure 2. SHAP value of group 0

5. まとめ

本稿では、新生児の心電図データから個人によって心電図に違いがあるのかを検証するために、k-means++法を用いてクラスタリングを行った。また、SHAP を用いてクラスタリング結果の理由付けを行った。しかし、理由付けに不明瞭な点があるため、パラメータの改善や、画像数を増やして検証を続ける。

参考文献

[1] 福崎, 他, 特集: ヒトの発達科学, (1998)12 巻 7 号, p.79-88
 [2] Craft N, Schwartz JB: (1995) Am J Physiol 268:H 1441-H 1452,
 [3] Lundberg, S. M., & Lee, S. I. (2017). Advances in neural information processing systems, 30.
 [4] GitHub – shap/shap
<https://github.com/shap/shap>, (2023.09.27)