

小規模な心筋細胞画像のテクスチャ特徴量による 心筋症状態分類予測と判定要因分析

○森 雅也¹, 小山 裕², 大前 佑斗¹, 原 一之¹, 豊谷 純¹, 奥村 恭男³, 羽尾 裕之²

¹ 日本大学 生産工学部

² 日本大学 医学部 病態病理学系 人体病理学分野

³ 日本大学 医学部 内科学系 循環器内科学分野

1. まえがき

特発性心筋症は根本的な治療法のない難治性心疾患で、その診断には心内膜下心筋生検による病理診断が重要な役割を果たす。一方で、心臓病理を専門とする病理医はごくわずかで、診断精度の確保が課題となっている。そのため、機械学習を用いた心臓病理診断を可能とする分類モデルが必要とされている。この実現には、個人差の影響が最小限に抑えられた、心筋細胞画像が持つ様々な心筋症の固有の情報を捉えることが可能な特徴量が不可欠である。そこで本研究では、心筋症の状態分類を対象に、心筋細胞画像から心筋症診断に有効である可能性の高い特徴量の生成・抽出を行う。また、これらの実験で使用されたデータセットにて、試験的な心筋症状態分類予測モデルの構築を行う。

2. 実験概要

本研究では、心筋症の状態として、典型例・境界例・正常例の3つを対象とする。3つの状態における心筋細胞画像を図1に示す。これらは256階調のグレー画像であり、1つの状態につき2人から、それぞれ10枚ずつ収集された。つまり、被験者は全員で6人であり、収集された心筋細胞画像は全部で60枚である。ここで、典型例の被験者をSub A, Sub B, 境界例の被験者をSub C, Sub D, 正常例の被験者をSub E, Sub Fと定義する。これらの心筋細胞画像に対して、(1) テクスチャ特徴量への変換, (2) 特徴量選択とテクスチャ特徴量の生成パラメータの探索, (3) 選択された特徴集合の次元圧縮, (4) 分類モデルの構築, の4つを実施した。

初めに、心筋細胞画像からそれぞれの状態が持つ固有の特徴を捉えやすくするために、画像の質感や模様を定量化したテクスチャ特徴量への変換を行なった。本研究では、テクスチャ解析手法として、濃淡ヒストグラム(FOF)・濃淡差分統計量(GLDS)・同時生起行列(GLCM)・ランレンクス行列(GLRLM)・フーリエパワースペクトルの角度方向分布(FPS-ADF)と動径方向分布(FPS-RDF)を採用した。また、テクスチャ特徴量は、これらの出力値を必要に応じて確率に変換した値の統計量とした。具体的には、FOF・GLDS・FPS-ADF・FPS-RDFでは、平均・コントラスト・分散・歪度・尖度・エネルギー・エントロピーを算出した。GLCMでは、エネルギー・コントラスト・相関・逆分散・逆差分モーメント・エントロピーを算出した。GLRLMでは、Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-Uniformity (GLN), Run Length Non-Uniformity (RLN), Run Percentage (RP) を算出した。この時、GLCMとGLRLMに関しては、PythonのオープンソースパッケージであるPyRadiomics¹⁾を使用した。以上により、全39種の特徴量を獲得した。これらは特徴量ごとにスケールが異なるため、60サンプル全てにおいて、平均が0、標準偏差が1の標準化を適用した。

テクスチャ特徴量は、医療画像と機械学習による多くの研究で使用されており、その有効性が広く報告されている^{2,3)}。しかし、有効性の高い特徴量は、対象とするデータセットや問題により異なる。本研究では、3つの状態を明確に判別することが可能であり、かつ個人差の影響

Cardiomyopathy State Classification and Determinant Analysis Using
Texture Features from Small Sample Size Cardiomyopathyte Images

Masaya MORI, Yutaka KOYAMA, Yuto OMAE, Kazuyuki HARA,
Jun TOYOTANI, Yasuo OKUMURA and Hiroyuki HAO

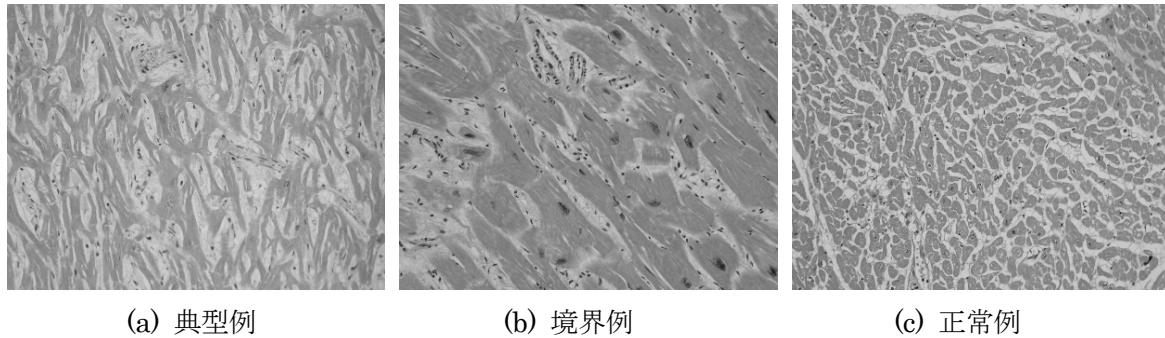


図1 心筋症の3つの状態における心筋細胞画像

響が小さい特徴量を、有効性が高いと判断する。このような特徴量を獲得するために、クラス内分散クラス間分散比を用いた特徴量選択を実施した。これは、全サンプルの平均ベクトルと各状態の平均ベクトルとの距離を、各状態におけるサンプルの分散で除算した値を評価値とする手法である。このとき、各状態間の距離が十分に離れており、かつ各状態のサンプルのばらつきが小さい特徴量が望ましいため、評価値は大きいほど良いと判断される。具体的な適用方法は、分類の対象となる状態の組み合わせごとに参考となる特徴は異なると考えられるため、(典型例, 境界例), (典型例, 正常例), (境界例, 正常例)の3つの分類問題に細分化した。これらの問題において、それぞれの評価値が最大となる上位3つの特徴量を選択した。このとき、特徴量の評価は1次元ずつ行い、個人差の影響が小さい特徴量を獲得するために、Sub AからSub Fまでの全ての被験者を対象とした。最終的に、選択された9つの特徴量の重複を除外した特徴集合が、3つの状態分類において最良な組み合わせと考えられるため、これを採用した。また、テキストチャ特徴量の生成にはいくつかのパラメータが存在し、これは特徴量の良し悪しに大きな影響を与える。そこで、ハイパーパラメータチューニングとして、各分類問題における上位3つの評価値の合計が最大となるパラメータの探索を行った。パラメータは、FOF・GLDS・GLCM・GLRLMに関しては、階調値を[2, 4, 8, 16, 32, 64, 128, 256]の範囲で探索した。またGLDSは、階調値に加えて対象ピクセルとの距離を[1, 2, 3, 4, 5]の範囲で探索した。FPS-ADFとFPS-RDFに関しては、微小計測幅を[2, 4, 6, 8, 10]の範囲で探索した。これらの探索はoptuna⁴⁾によるベイズ最適化により実施した。

本研究の対象画像はサンプルサイズが60と小さいため、特徴量の次元数が高い場合、過学

習という分類モデルの汎化性能が低下する問題が生じる可能性がある。また、抽出した特徴量の有効性を視覚的に評価するために、ここでは選択された特徴集合の次元圧縮を実施する。本研究では、高次元空間におけるクラス間の距離情報の損失を最小限に抑えるために、教師あり次元圧縮法の1つであるフィッシャーの線形判別分析を採用した。また、過学習の抑制だけでなく可視化することも目的としているため、第二主成分までの出力とした。これにより、特徴量が2次元、サンプルサイズが60のデータセットに変換された。

最後に、試験的な心筋症状態分類予測モデルの構築について説明する。ここでは、過学習の抑制や要因分析の容易化のために、モデル容量の小さいLinear Support Vector Machine (LSVM)を採用した。LSVMはscikit-learnにより実装し、パラメータはデフォルト値を採用した。また、手元のデータセットに対する分類性能に加えて、個人差の影響度も確認するために、モデル構築に使用するTrainデータをSub A・Sub C・Sub Eのサンプル30件とし、性能評価に使用するTestデータをSub B・Sub D・Sub Fのサンプル30件とした。

3. 結果と考察

初めに、特徴量選択手法により選出されたテキストチャ特徴量と、選択された特徴量におけるサンプルの平均と標準偏差を確認する。これらを表1に示す。ここでは、全体としてFOF・GLCM・FPS-RDFが選出された。このときの生成パラメータは、FOFの階調値が32、GLCMの階調値が8、FPS-RDFの微小計測幅が4であった。それぞれの分類問題に着目すると、(典型例, 境界例)では、FOFの平均と尖度とコントラストが選出された。また、境界例と比較して、典型例はFOFの平均とコントラストのサンプル平均が大きく、FOFの尖度のサンプル

表 1 評価値が上位 3 位のテクスチャ特徴量

分類問題	テクスチャ特徴量	平均 ± 標準偏差
(典型例, 境界例)	FOF : 平均	(152.61 ± 4.43, 146.60 ± 4.30)
	FOF : 尖度	(2.77 ± 0.36, 3.47 ± 0.68)
	FOF : コントラスト	(23908 ± 1285, 22113 ± 1275)
(典型例, 正常例)	GLCM : 相関	(0.76 ± 0.01, 0.69 ± 0.02)
	FPS-RDF : 歪度	(1.49 ± 0.06, 1.25 ± 0.06)
	FPS-RDF : 尖度	(4.63 ± 0.21, 3.92 ± 0.19)
(境界例, 正常例)	GLCM : 相関	(0.76 ± 0.03, 0.69 ± 0.02)
	FPS-RDF : 歪度	(1.51 ± 0.14, 1.25 ± 0.06)
	FPS-RDF : コントラスト	(1405 ± 155, 1717 ± 109)

平均が小さいことが確認された。FOFは、対象画像の階調値の頻度を全体の頻度で除算した、階調値の出現確率分布を表している。そのため、典型例は心筋細胞画像の階調値が比較的高い傾向が、境界例は特定の階調値に比較集中する傾向があると考えられる。(典型例, 正常例)では、GLCMの相関・FPS-RDFの歪度と尖度が選出された。また、正常例と比較して、典型例は全特徴量のサンプル平均が大きいことが確認された。GLCMは、隣接するピクセル間における階調値の同時出現確率を表している。これの相関が高くなる条件は、階調値の最小値付近と最大値付近との同時出現確率が高く、反対に、近い階調値の同時出現確率が低いことである。またFPS-RDFは、画像の全方向において、低周波成分が多い場合に右裾の長い分布となり、高周波成分が多い場合に左裾の長い分布となる。そのため、正常例の心筋細胞画像は、隣接するピクセル間における階調値が比較的緩やかに変化し、さらに全方向において比較的細かい濃淡周期をもつ傾向があると考えられる。(境界例, 正常例)では、GLCMの相関・FPS-RDFの歪度とコントラストが選出された。また、正常例と比較して、境界例でも全特徴量のサンプル平均が大きいことが確認された。そのため、(典型例, 正常例)と同じ傾向があると考えられる。

次に、LSVMの推定結果と識別境界の可視化から、本研究で獲得されたテクスチャ特徴量が、個人差の影響を抑えつつ各状態の明瞭な判別を可能にできるかの確認を行う。推論結果は、混同行列とF値のマクロ平均により確かめる。混同行列とは、分類モデルの真陽性・真陰性・偽陽性・偽陰性の4つをまとめたものである。

表 2 Train データにおける混同行列

		予測値		
		典型例	境界例	正常例
正解値	典型例	10	0	0
	境界例	3	7	0
	正常例	0	0	10

表 3 Test データにおける混同行列

		予測値		
		典型例	境界例	正常例
正解値	典型例	10	0	0
	境界例	1	9	0
	正常例	0	0	10

F値のマクロ平均とは、多クラスを対象とした分類モデルの推定性能を[0, 1]の範囲で評価する指標である。これは、評価値が1に近いほど推定性能が高いと判断することが可能である。Trainデータにおける混同行列を表2に、Testデータにおける混同行列を表3に示す。表2に着目すると、典型例と正常例に関しては、明確な判別ができていたことが確認された。一方で、境界例に関しては、3サンプルにおいて典型例と誤った判断を下していることが確認された。このときのF値のマクロ平均は0.90であった。続いて表3に着目すると、ここでも典型例と正常例に関しては、明確な判別が可能であること

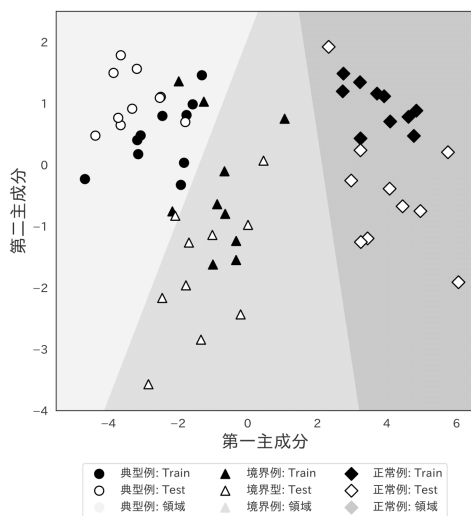


図2 2次元特徴量空間における識別境界

が確認された。また、境界例に関しては、典型例と誤分類したサンプルが1つであることが確認された。このときのF値のマクロ平均は0.97であった。続いて、次元削減により獲得した2次元特徴量空間と分類モデルの識別境界を確認する。これらを図2に示す。ここで、各状態におけるTrain・Testデータに着目すると、被験者が異なる場合でも、各状態のサンプルが特定の領域に集中していることが確認された。また、各状態間の距離が離れており、特に正常例においては、識別境界付近にサンプルが存在せず明瞭に分類できていることが確認された。

以上から、本研究で獲得されたテクスチャ特徴量には、個人差の影響を最小限に抑制できる可能性があることが示された。また、第三者に対する心筋症状態分類予測の実現可能性が示された。

4. まとめ

本研究では、機械学習による心臓病理診断の実現に向けて、3種の心筋症を対象に、テクスチャ特徴への変換抽出と試験的な状態分類予測モデルによる有効性検証を行った。これは、心筋細胞画像からテクスチャ特徴量を生成し、特徴量選択・テクスチャ特徴量の生成パラメータ探索・次元圧縮を適用した後に、LSVMにより検証された。その結果、状態分類に最適なテクスチャ特徴として、FOFの平均と尖度とコントラスト・GLCMの相関・FPS-RDFの歪度と尖度とコントラストの7つが選出された。またLSVMによる検証では、TrainデータのF値のマクロ平均で0.90、TestデータのF値のマ

クロ平均で0.97が達成され、識別境界の可視化から、個人差の影響が小さく、3つの状態を明瞭に分類可能な特徴量であることが確認された。以上から、3種の心筋症診断において、本研究で変換抽出されたテクスチャ特徴量は心筋症診断に有効である可能性が示された。

今後は、サンプルサイズを増やすことで、テクスチャ特徴量のより厳密な有効性の検証を行う。また、本研究は心筋症診断に有効な可能性のある特徴量の探索に主軸を置いており、心筋症状態分類予測モデルの構築に関しては、データセットの関係上、手元にある既知のデータを使用したため、結果として予測性能が高く評価されている可能性がある。今後は、第三者の未知のデータによる検証を行い、分類モデルの実現可能性について詳細な調査を行う。

Ethical Statement

本研究は、日本大学医学部附属板橋病院倫理審査委員会 RK-210914-18（研究代表者：羽尾裕之）の承認を得た上で実施されている。

謝辞

本研究は、科学研究費助成事業 基盤研究（C）23K11310 の支援を受けた。また、電気通信大学 庄野逸先生には、テクスチャ特徴量の解説資料をご提供頂き、本研究の遂行に多大なるご協力を頂いた。

参考文献

- 1) van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J. C., Pieper, S., & Aerts, H. J. W. L, “Computational radiomics system to decode the radiographic phenotype”, *Cancer research*, 77(21), 2017, e.104-107.
- 2) Liu, Q., Jiang, P., Jiang, Y., Ge, H., Li, S., Jin, H., & Li, Y., “Prediction of aneurysm stability using a machine learning model based on PyRadiomics-derived morphological features”, *Stroke*, 50(9), 2019, pp.2314-2321.
- 3) Xv, Y., Lv, F., Guo, H., Zhou, X., Tan, H., Xiao, M., & Zheng, Y., “Machine learning-based CT radiomics approach for predicting WHO/ISUP nuclear grade of clear cell renal cell carcinoma: an exploratory and comparative study”, *Insights into imaging*, 12(1), 2021, pp.1-14.
- 4) Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M., “Optuna: A next-generation hyperparameter optimization framework”, *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp.2623-2631.