

画像認識による縫合画像の識別

池末千峰奈¹⁾ 橋本大輔²⁾ 松村和季²⁾ 荒平高章¹⁾

概要: 本研究では、医学部教育における縫合技能評価を支援するため、縫合画像を入力とした自動評価手法を提案する。医師により5段階評価された縫合画像を用い、畳み込みニューラルネットワーク (CNN) による分類を行った。縫合間隔 (interval) に着目して性能を検証した結果、テストデータにおいて約80%の分類精度を得た。本手法は縫合技能評価の客観化への可能性を示す。

キーワード: 2170301 医療・福祉支援、2150205 画像認識・理解、2170105 教育支援

Automatic Classification of Suture Images Using Convolutional Neural Networks

CHIHONA IKEMATSU¹⁾ DAISUKE HASHIMOTO²⁾ KAZUKI MATSUMURA²⁾
TAKA AKI ARAHIRA¹⁾

Abstract: This study proposes an automatic evaluation method for suture skill assessment in medical education. Using five-level surgeon-labeled suture images, a convolutional neural network (CNN) was trained to classify interval quality. The proposed method achieved approximately 80% accuracy on the test dataset, demonstrating the feasibility of objective suture evaluation based solely on image information.

Keywords: 2060105 Database design, 2120200 Network security, 2170301 Medical and welfare support

1. はじめに

外科手技の一つである縫合は、医師に求められる基本的かつ重要な技能であり、医学部教育や臨床研修において反復的な訓練が行われている。縫合技能の習得には、縫合間隔の正確さ、縫合の強さ、創縁の幅の均一性など複数の要素が関係しており、これらを総合的に評価することが求められる。

しかしながら、現在の縫合技能評価は主に指導医による目視評価や主観的判断に依存しており、評価者の経験や負担に大きく左右されるという課題がある。また、評価基準が定量化されにくく、再現性に乏しいことも問題として指摘されている。さらに、多数の学生や研修医を対象とする教育現場では、指導医一人あたりの評価負担が大きく、個々の学習者に対して十分なフィードバックを行うことが困難である。

このような背景から、客観的かつ効率的に縫合技能を評価する支援システムの構築が求められている。

本研究では、医学部教育における縫合トレーニングを支援することを目的とし、縫合画像を入力とした画像認識による縫合評価手法を提案する。医師により5段階で評価された縫合サンプル画像を教師データとして用い、畳み込みニューラルネットワーク (CNN) による自動分類を試みた。本研究では特に縫合間隔 (interval) の評価に着目し、画像情報のみから評価段階のうちから1と5を推定できるかを検証する。

2. 理論

畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) は、画像認識分野において高い性能を示す深層学習モデルである。CNN は畳み込み層およびプーリング層を用いて画像の局所的特徴を抽出し、階層的に抽象化された特徴表現を学習する。本研究では、縫合画像に含まれる縫合間隔や糸配置のパターンを特徴量として自動抽出するためにCNNを用いた。

3. 提案手法

3.1 データ収集

本研究では、医師より提供された縫合サンプルを対象と

1 九州情報大学
Kyushu Institute of Information Sciences, Dazaifu, Fukuoka 818-0117, Japan
2 関西医科大学
Kansai Medical University, Hirakata, Osaka 573-1010, Japan

して画像データを収集した。縫合サンプルは縫合練習後の結果を撮影したものであり、撮影は同一環境下で行われた。

撮影には自身の所持するスマートフォン (google pixel 8) を用い、縫合部全体が画面内に収まるように撮影角度および距離を統一した。背景や照明条件の影響を極力抑えるため、撮影時の環境は可能な限り一定となるよう配慮した。収集した画像は JPEG 形式で保存し、本研究で提案する画像認識手法の入力データとして使用した。

3.2 教師ラベル

本研究で使用した画像データには、医師による縫合評価結果を教師ラベルとして付与した。評価項目は、縫合間隔 (interval)、縫合幅 (width)、縫合の強さ (strength) の3項目であり、それぞれ5段階評価 (1~5) で判定されている。

本研究では、まず縫合間隔 (interval) の評価に着目し、画像から評価段階を推定する分類問題として扱った。各画像は、医師の評価結果に基づき対応するクラスラベルが割り当てられている。

このように、専門知識を有する医師の評価を教師データとして用いることで、教育現場に即した実用的な分類モデルの構築を目指した。

3.3 前処理

入力画像に対して、学習の安定化および計算効率の向上を目的とした前処理を行った。

まず、各画像を固定サイズである 100×100 ピクセルにリサイズし、入力サイズを統一した。次に、画素値を 0 から 1 の範囲に正規化することで、ニューラルネットワークの学習が行いやすくなるよう調整した。

これらの前処理により、撮影条件のばらつきによる影響を軽減し、CNN による特徴抽出が効果的に行われるようにした。

3.4 データ拡張

本研究では、データ数の不足および過学習の抑制を目的としてデータ拡張を行った。医療画像においては撮影角度や照明条件の違いにより見え方が変化するため、実際の利用環境を想定した拡張処理が重要である。

具体的には、入力画像に対して以下の処理を施した。

- ・左右反転 (horizontal flip)
- ・明度およびコントラストのランダム変化

これらの処理により、学習データの多様性を人工的に増加させ、モデルの汎化性能向上を図った。

3.5 CNN モデル構造

本研究では、縫合画像の特徴抽出および分類を行うために畳み込みニューラルネットワーク (CNN) を用いた。モデルは Sequential 構造を採用し、以下の構成とした。

1. 畳み込み層 (Conv2D) + 活性化関数 ReLU
2. プーリング層 (MaxPooling)
3. 上記の構成を複数回繰り返し、特徴マップを抽出
4. Flatten 層による次元ベクトル化
5. 全結合層 (Dense)
6. Dropout 層による過学習抑制
7. 出力層 (Softmax) による 5 クラス分類

畳み込み層では画像の局所的特徴 (縫合間隔のパターンや糸の配置) を抽出し、プーリング層により空間情報を圧縮することで計算量を削減した。

また、Dropout 層を導入することで過学習を防止し、未知データへの汎化性能向上を図った。

4. 実験

4.1 実験設定

本研究では、縫合間隔 (interval) の評価 1 および評価 5 を対象とした二値分類実験を行った。評価 1 および評価 5 の画像のみを抽出し、元画像単位で訓練データ・検証データ・テストデータに分割した。データリークを防止するため、拡張処理は訓練データのみに適用した。

モデルには MobileNetV2 を用いた転移学習モデルを採用した。ImageNet で事前学習された重みを利用し、特徴抽出層は凍結した状態で学習を行った。出力層はシグモイド関数を用いた二値分類構造とした。

損失関数にはバイナリクロスエントロピーを使用し、最適化手法には Adam を採用した。バッチサイズは 16、エポック数は 30 とした。さらに、検証精度が一定期間改善しない場合に学習を停止する Early Stopping および学習率減衰 (ReduceLRonPlateau) を導入し、過学習の抑制および学習安定化を図った。

4.2 データ分布

本実験では、縫合間隔評価 1 および評価 5 のデータのみを使用した。最終的なテストデータ数は 39 枚であり、評価 1 が 21 枚、評価 5 が 18 枚であった。

本実験は両極端の評価 (1 と 5) に限定した二値分類であり、識別が比較的明確なケースを対象としている。中間評価を含む多クラス分類とは難易度が異なる点に留意する必要がある。

4.3 結果

学習の結果、以下の精度を得た。

- ・分類精度 (train best) : 0.9919
- ・検証精度 (val best) : 0.9677
- ・テスト精度 (test) : 0.9231

テストデータ 39 枚中 36 枚を正しく分類し、誤分類は 3

枚であった。

混同行列は以下の通りである。

$$\begin{bmatrix} 19 & 2 \\ 1 & 17 \end{bmatrix}$$

すなわち、評価 1 を正しく判定した件数 (TN) は 19 件、評価 5 を正しく判定した件数 (TP) は 17 件であった。誤分類は、評価 1 を 5 と判定したケースが 2 件、評価 5 を 1 と判定したケースが 1 件であった。

classification report の結果は以下の通りである。

- ・評価 1 : Precision 0.9500、Recall 0.9048、F1-score 0.9268
- ・評価 5 : Precision 0.8947、Recall 0.9444、F1-score 0.9189

以上より、縫合間隔評価の両極端においては高い識別性能が得られた。

4.4 失敗例の分析

誤分類は 3 件のみであり、その多くは縫合間隔が境界的である画像であった。具体的には、評価 1 と評価 5 の中間の特徴を有する画像や、照明条件により縫合糸の視認性が低下している画像であった。

これらの結果から、極端な評価間においても完全な識別は困難であり、撮影条件および縫合のばらつきが分類結果に影響を与える可能性が示唆された。

5. 考察

5.1 精度の解釈

本研究において、縫合間隔評価 1 と 5 の二値分類に対するテスト精度は 0.9231 であった。これは未知の縫合画像に対して約 92% の確率で正しい評価を推定できたことを示している。

縫合評価は本来、医師の主観的判断に基づくものであり、完全に客観化することは容易ではない。そのような状況下において、画像のみから高精度で識別できたことは、深層学習による自動評価の有効性を示す結果である。

特に、precision および recall が両クラスで 0.89 以上を示していることから、誤検出および見逃しの両面において安定した性能を有していると考えられる。

5.2 本研究の限界

本研究には以下の制約がある。

第一に、本実験は評価 1 と評価 5 のみに限定した二値分類であり、中間評価を含む多クラス分類は実施していない点である。実際の教育現場では 5 段階評価が用いられるため、今後は多クラス分類への拡張が必要である。

第二に、データ数が限定的であり、特定の撮影環境に依存している可能性がある。異なる環境や施設においても同様の性能が得られるか検証する必要がある。

5.3 応用可能性

本研究の成果は、医学部教育における補助的評価システムとして活用できる可能性がある。学習者が縫合練習後に画像を撮影し、即時に評価結果を得ることで、自己フィードバックの促進および指導医の負担軽減が期待される。

本手法は医師の評価を代替するものではなく、補助的ツールとして位置付けられるが、客観的指標の一つとして教育支援に寄与できると考えられる。

5.4 今後の課題

今後の課題として以下が挙げられる。

- ・5 段階多クラス分類への拡張
- ・データ数の拡充および多施設データによる検証
- ・interval・width・strength を統合した多項目同時評価モデルの構築
- ・動画データを用いた動作解析の導入

特に、実際の縫合技能は動作過程も重要であるため、将来的には時系列情報を取り入れた評価モデルの検討が必要である。

6. まとめ

本研究では、縫合技能評価の客観化を目的として、縫合画像を対象とした深層学習による自動分類手法を提案した。評価 1 および評価 5 の二値分類実験において、テスト精度 0.9231 を達成し、高い識別性能を示した。

本結果は、画像認識技術が縫合技能評価の補助ツールとして活用可能であることを示す基礎的成果である。

今後は多クラス分類への拡張および多項目統合評価モデルの構築を目指す。また、動画解析や多視点撮影を取り入れることで、より実践的な技能評価システムの実現を目指す。

最終的には、医学教育現場において学習者が自己評価を行える支援環境を構築することを目標とする。

参考文献

- [1] 平川智喜. (2022). 深層学習を用いたインプラントの画像識別. 卒業論文.