

変分オートエンコーダを用いた個人特徴の分離による歩容映像に基づく疾患推定 Disease Estimation Based on Gait image by Individual Feature Separation Using Variational Autoencoder

古川 栞¹⁾ 武村 紀子¹⁾
Shiori Furukawa Noriko Takemura

1 はじめに*

高齢化社会に伴い、特に患者数が多い症例として歩行が困難になる歩行障害がある。歩行障害は行動を制限されるだけでなく、歩行時間の減少による物忘れの進行や意欲の減少といった精神的な影響を患者に及ぼす。また、社会的な観点から、歩行障害を有する患者の介護に労力を要することも問題となっている。

歩行障害を有する代表的な疾患として、腰部椎管狭窄症 (lumber canal stenosis, 以下, LCS) や頸椎症性脊髄症 (cervical spondylotic myelopathy, 以下, CSM)、パーキンソン病、末梢動脈疾患、脳血管疾患などが挙げられる。これらの疾患は専門とする診療科が多岐に渡り、また、バイオマーカー等の簡便なスクリーニング検査法が欠如しているため、正しい診断結果を受けるまでにかなりの時間を要する場合がある。実際、頸椎症患者の 43% は当初は他疾患と診断されることや、歩行障害のある患者が適切な診断に至るまで平均 5.2 人の医師を受診している現状がある [1]。

そこで本研究では、人の歩き方の特徴からこれらの疾患を自動で推定することを目指す。

これまで歩行障害を有する疾患における歩容分析に関する研究がいくつか存在する [2, 3, 4, 5]。Tahir ら [3] は、モーションキャプチャシステムおよび床反力計を用いて、関節角度、ストライド長、歩行時の床反力等の特徴を抽出し、機械学習モデルによりパーキンソン病患者の識別を行っている。しかし、この手法では、専門的知識を必要とする高価なセンサを使用しているため、実用的な診断やスクリーニング検査には不向きである。

一方、Kidziński ら [4] は、1 台のカメラで撮影した歩容映像を用いて、歩行速度、ケイデンス、膝関節角度等を推定している。歩く姿をカメラで撮影するだけで歩容特徴が推定可能であり実用性は高いが、この手法では、大まかな骨格モデルに基づいて上記の特徴を推定しており、首や腰の微妙な曲がり具合など、疾患との関係が深い多くの情報が欠落している。さらに関節点の推定誤差や誤検出が疾患推定の精度低下をもたらす恐れがある。そこで本研究では、骨格特徴に基づくモデルベースの手法ではなく、シルエット特徴に基づくアピアランスベースの手法を採用し、歩容映像から疾患推定を行う。

アピアランスベースの手法は画像から直接疾患推定を行うため、疾患に関する情報の欠落が少ない。しかし同時に、髪型や体型といった個人の特徴も画像に含まれるため、これらの個人特徴が疾患推定性能に影響する可能性がある。そのため、いかにして個人によらない疾患の特徴を抽出するかが課題となる。本研究では、変分オートエンコーダ (Variational Auto Encoder, 以下 VAE) をベースとした Disentangled Representation

Learning (以下 DRL) により、この問題に対処する。Disentangled な表現とは、潜在空間中の各次元が被写体の因子 (色, 形, 大きさ等) ごとに分かれている状態を表しており、潜在空間において疾患を表す特徴と個人を表す特徴を分離させることで、個人によらない疾患特徴を抽出する。一般に、VAE をベースとした DRL は、因子を指定可能な画像生成問題、例えば、髪の色や表情を指定した人物画像生成などに用いられるが [6]、ここでは、識別問題において個人差の影響を抑制させることを目的とする。

2 提案手法

本研究では、歩容映像から作成した平均シルエット画像 (Gait Energy Image, 以下 GEI) を用いて、VAE をベースとした DRL モデルにより疾患の推定を行う。以下に提案手法の詳細を述べる。

2.1 歩容特徴

疾患推定モデルへ入力する GEI の生成手順を図 1 に示す。まず、カメラで撮影した歩容映像 (RGB 画像列) に対してグラフ転移学習を用いた人物領域分割法 [7] を適用し、シルエット画像列を得る。次に、シルエット画像列に対し身長で正規化を行い、128 × 88 画素の正規化シルエット画像列を生成する。最後に正規化シルエット画像列 40 枚分に対して各画素値の平均を取ることで、1 枚の GEI (128 × 88 画素) を生成する。GEI を生成する際は 1 歩行周期分の画像列の平均を取るのが一般的であるが、本研究で対象としている歩行障害を有する患者は、歩行が安定せず正しい歩行周期が推定できないケースがあるため、ここでは実験的に 40 フレーム分の画像列から GEI を生成することとした。

GEI は首や腰の曲がり具合といった静的特徴と手足の振り方といった動的特徴を 1 枚の画像で表現しており、歩容解析や歩容認証における様々な研究で用いられている有効な歩容特徴である。しかし一方で、図 2 のように体型や髪型といった個人の特徴も多く含むため、歩行を解析する際はこのような個人差の影響を考慮する必要がある。

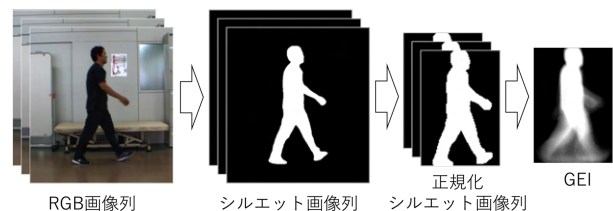


図 1: GEI 生成手順

1) 九州工業大学 Kyusyu Institute of Technology

* 本論文は画像の認識・理解シンポジウム MIRU2023 においてコンセプト論文として発表した内容に基づく



図 2: 個人差のある GEI

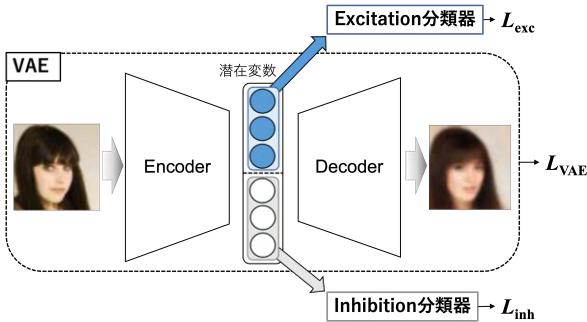


図 3: DRL モデルの概要

2.2 DRL モデル

本研究では, Guided-VAE[8] に基づき, VAE の潜在空間において特徴の分離を行う. 本 DRL モデルは図 3 に示すように, VAE モデル, Excitation 分類器, Inhibition 分類器の 3 つのネットワークにより構成される. ただし, 各ネットワーク構造の詳細は文献 [8] と同様のものを用いるものとする.

VAE モデル 入力画像を一度圧縮させた後, 入力画像と同じ画像を再構築するネットワークであり, 損失関数 L_{VAE} は入力画像と出力画像の再構築誤差 (平均二乗誤差) および潜在変数分布と正規分布との Kullback-Leibler divergence を足し合わせたものとする.

Excitation 分類器 ある潜在変数が特定の特徴情報を持つように学習する際に使用する分類器で, 損失関数 L_{exc} は Binary Cross Entropy (BCE) とする.

Inhibition 分類器 残りの潜在変数が特定の特徴情報を持たないように学習する際に使用する分類器で, 損失関数 L_{inh} は L_{exc} と同様, BCE とする.

本 DRL モデルの学習手順を以下に示す. この学習手順を繰り返すことで, 特定の特徴情報を持つ潜在変数と持たない潜在変数に分離することができる.

1. 損失関数を $L_{VAE} + \alpha L_{exc}$ として, VAE 及び Excitation 分類器のパラメータを学習させる. ただし, α は L_{VAE} と L_{exc} の足し合わせる際の重みである. ここでは, Excitation 分類器が特徴ラベルを正しく分類できるように, また, VAE が正しい画像を再構築でき, Excitation 分類器が正しく分類できるように特徴が潜在変数に含まれるように学習される.
2. 損失関数を L_{inh} として, Inhibition 分類器を学習させる. ここでは, Inhibition 分類器が特徴ラベルを正しく分類できるように学習される.
3. 特徴ラベルの値をすべて均一 (ラベル値 = $1/\text{クラス数}$) にし, 損失関数を L_{inh} として VAE を学習させる. ここでは, 潜在変数が特定の特徴情報を持た

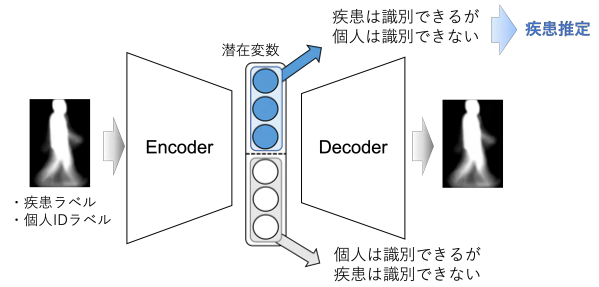


図 4: 提案手法のフレームワーク

	個人ID				
	1	2	3	4	5
教師ラベル = [0	0	1	0	0
均一ラベル = [0.2	0.2	0.2	0.2	0.2

図 5: 個人 ID ラベルの One-hot ベクトル表現

ないように学習される.

2.3 疾患推定手法

前節で述べた DRL モデルを用いて, 個人特徴を考慮した疾患推定を行う. 本手法のフレームワークを図 4 に示す. 入力 GEI には疾患の有無と個人 ID のラベルが付与されている. 疾患特徴, 個人特徴それぞれについて, 前節で述べた手順を繰り返すことで, 疾患は識別できるが個人は識別できない特徴と, 個人は識別できるが疾患は識別できない特徴とに分離する. 前者の特徴, つまり, 個人の特徴を含まない疾患特徴を用いることで, 個人差を考慮した疾患推定が可能となる.

疾患特徴の分離 図 4 のように潜在変数を 2 つに分割し, 上半分の潜在変数に対して Excitation 分類器を, 下半分の潜在変数に対して Inhibition 分類器をそれぞれ適用し, 疾患ラベルを用いて学習する. ただし, 疾患ラベル数は 2 であるため, DRL モデルの学習手順 (3) の均一ラベルの値は 0.5 とする. 以上より, 上半分の潜在変数には疾患特徴が含まれるように, 下半分の潜在変数は疾患特徴を含まないように学習される.

個人特徴の分離 疾患特徴の場合と同様に, 上半分の潜在変数に対して Inhibition 分類器を, 下半分の潜在変数に対して Excitation 分類器をそれぞれ適用し, 個人 ID ラベルを用いて学習する. ただし, 個人 ID ラベルは図 5 のように One-hot ベクトル表現を用いるものとし, DRL モデルの学習手順 (3) の均一ラベルの値は 1 を ID 数で割った値とする (ID 数=5 の場合, 均一ラベル値=0.2). 以上より, 上半分の潜在変数には個人特徴が含まれないように, 下半分の潜在変数には個人特徴が含まれるように学習される.

3 評価実験

歩行障害を有する患者および健常者の歩容映像を収集し, 評価実験を行った.

表 1: 被験者のデータ数内訳

疾患	被験者数	GEI 枚数
LCS	142	49,505
CSM	133	28,053
LCS または CSM	223	67,398
健常者	332	28,817



図 6: 歩行映像の例

3.1 データセット

一般的な単眼 RGB カメラを用いて、LCS 患者、CSM 患者および健常者の 4m 分の歩行映像（画像サイズ：1288 × 964 画素、フレームレート：30fps）を撮影した。図 6 に撮影した歩容映像の一例を示す。LCS 患者 142 人、CSM 患者 133 人（うち 52 人は LCS、CSM どちらも罹患）、健常者 332 人に対し、1 人あたり 1~4 回、計 1086 回分の歩容映像を収集した。各歩容映像に対して、医師による診断ラベルおよび個人 ID ラベルが付与されている。40 フレーム分の画像列から GEI を生成する作業を、1 フレームずつずらして行うことで、約 96,000 枚の GEI を生成した。ただし、歩行速度の違いにより 1 回の歩容映像から生成される GEI の枚数が異なるため、1 回の歩容映像から生成する GEI の上限枚数を 200 とした。疾患ごとの被験者数、GEI 枚数を表 1 に示す。

3.2 比較手法

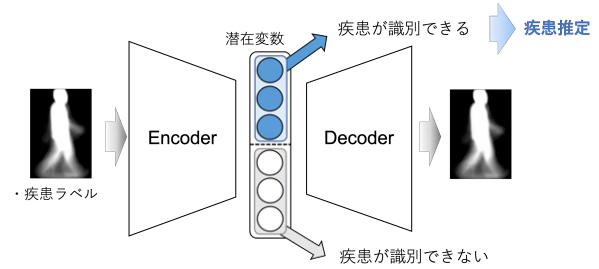
個人差を考慮した本疾患推定手法の有用性を示すために、以下の 2 つの比較手法を用いて性能評価を行う。

比較手法 1 提案手法において、疾患特徴の分離のみを行い、個人特徴の分離を行わない場合（図 7(a)）。つまり、上半分の潜在変数に対して Excitation 分類器を、下半分の潜在変数に対して Inhibition 分類器をそれぞれ適用し、疾患ラベルを用いて学習する。

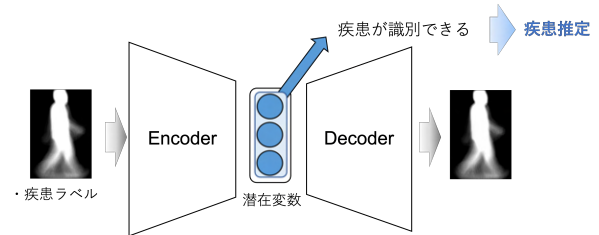
比較手法 2 潜在空間における特徴の分離は行わない場合（図 7(b)）。つまり、潜在変数に対して Excitation 分類器のみを適用し、疾患ラベルを用いて学習する。ただし、疾患推定に用いる潜在変数の次元数を他の手法と揃えるため、VAE の潜在変数は他手法の半分とする。また、DRL モデルを学習する際は、手順 (2)(3) は行わず、手順 (1) のみを行う。

3.3 評価手法

LCS 患者、CSM 患者、健常者の歩容データを用いて、以下の 3 つの疾患推定器を生成し、提案手法および比較



(a) 比較手法 1



(b) 比較手法 2

図 7: 比較手法のフレームワーク

手法についてそれぞれ性能評価を行う。

- **LCS 推定器:** LCS vs その他（CSM または健常）
- **CSM 推定器:** CSM vs その他（LCS または健常）
- **疾患推定器:** 疾患（LCS または CSM）vs 健常

性能評価の際は、歩容データベースにおいて被験者を 5 つのグループに分割し、交差検証を行う。5 グループのうち、1 グループをテストデータ、他の 1 グループを検証データ、残りの 3 グループを学習データとし、全てのグループが 1 回ずつテストデータになるように計 5 回検証を行い、5 回分の F1 score の平均値を用いて評価を行う。ただし、各識別クラスの学習データ数の偏りや、個人識別器におけるクラス数（個人 ID 数）の違いによる影響を考慮し、各グループに含まれる疾患あり/なしの被験者数が等しくなるようにデータを選出する。

学習の際のミニバッチサイズは 512 に設定し、検証データの F1 score が最大となるエポック数（最大 100 エポック）におけるモデルを用いてテストデータで性能評価を行う。 L_{VAE} と L_{exc} の重み α 、各ネットワークにおける初期学習率、Weight decay 等のハイパーパラメータについては、1 回目の交差検証時に検証データを用いて決定し、2 回目以降の交差検証においても 1 回目と同じ値を用いる。また、モデルの最適化には Adam を用いる。

3.4 実験結果

評価実験の結果を表 2 に示す。LCS 推定器と CSM 推定器では提案手法、比較手法 1、比較手法 2 の順に F1 score が高くなり、疾患推定器では全く逆の順番となった。

LCS 推定器と CSM 推定器においては、比較手法 2 よりも比較手法 1 の精度が高いことから、Inhibition 分類器を用いた潜在空間における疾患特徴の分離が有効に働いたと言える。さらに、比較手法 1 よりも提案手法の精度が高いことから、潜在空間における個人特徴の分離についてもその有効性を示せた。

一方疾患推定器については、全ての手法において F1

表 2: 実験結果 (平均 F1 score, 太字は最良値)

	LCS 推定器	CSM 推定器	疾患推定器
提案手法	0.863	0.822	0.952
比較手法 1	0.855	0.808	0.960
比較手法 2	0.851	0.789	0.963

score が 0.95 を上回る非常に簡単な問題設定となっており, 推定性能が飽和状態に陥っている. そのため, 最もネットワーク構造がシンプルでモデルの学習が容易な比較手法 2 の精度が高い結果となったと考えられる.

4 おわりに

本研究では, 歩容映像から抽出した GEI を用いて VAE により疾患推定を行う手法を提案した. GEI には歩容特徴以外に髪型や体型などの個人特徴が含まれており, これらの個人特徴は疾患推定の精度に影響すると考えられる. そこで提案手法では, DRL モデルを用いて VAE の潜在空間において疾患特徴と個人特徴を分離させることで, GEI の個人特徴を考慮した疾患推定を行った. 評価実験の結果, 提案手法では LCS 推定器および CSM 推定器の F1 score がそれぞれ 0.863, 0.822 となり, 個人特徴を考慮していない比較手法よりも高い推定性能を示した.

今後はパーキンソン病など, LCS や CSM 以外の疾患についても対象を広げ, 現在の 2 クラス分類のフレームワークから多クラス分類のフレームワークへと拡張を行う予定である.

参考文献

- [1] Jau-Ching Wu, Chin-Chu Ko, Yu-Shu Yen, Wen-Cheng Huang, Yu-Chun Chen, Laura Liu, Tsung-Hsi Tu, Su-Shun Lo, and Henrich Cheng. Epidemiology of cervical spondy-
- [2] Enas Abdulhay, N Arunkumar, Kumaravelu Narasimhan, Elamaran Vellaippan, and V Venkatraman. Gait and tremor investigation using machine learning techniques for the diagnosis of parkinson disease. *Future Generation Computer Systems*, Vol. 83, pp. 366–373, 2018.
- [3] Nooritawati Md Tahir and Hany Hazfiza Manap. Parkinson disease gait classification based on machine learning approach. *Journal of Applied Sciences(Faisalabad)*, Vol. 12, No. 2, pp. 180–185, 2012.
- [4] Łukasz Kidziński, Bryan Yang, Jennifer L Hicks, Apoorva Rajagopal, Scott L Delp, and Michael H Schwartz. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature communications*, Vol. 11, No. 1, p. 4054, 2020.
- [5] Trong-Nguyen Nguyen, Huu-Hung Huynh, and Jean Meunier. Skeleton-based abnormal gait detection. *Sensors*, Vol. 16, No. 11, p. 1792, 2016.
- [6] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. Disentangled controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5154–5163, 2020.
- [7] Ke Gong, Yiming Gao, Xiaodan Liang, Xiaohui Shen, Meng Wang, and Liang Lin. Graphonomy: Universal human parsing via graph transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7450–7459, 2019.
- [8] Zheng Ding, Yifan Xu, Weijian Xu, Gaurav Parmar, Yang Yang, Max Welling, and Zhuowen Tu. Guided variational autoencoder for disentanglement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7920–7929, 2020.