

# Deep Learning を用いた画像解析による 足部セルフチェックシステムの開発

新潟医療福祉大学 小 宮 諒  
(共同研究者) 第一工科大学 竹 下 康 文  
同 松 田 翔 太  
令和健康科学大学 中 井 雄 貴  
新潟医療福祉大学 江 玉 睦 明

## Development of a Foot Morphology Self-Check System by Image Analysis Using Deep Learning

by

Makoto Komiya, Mutsuaki Edama  
*Niigata University of Health and Welfare*  
Yasufumi Takeshita, Shota Matsuda  
*Daiichi Institute of Technology*  
Yuki Nakai  
*Reiwa Health Sciences University*

### ABSTRACT

Foot posture assessment is recognized as a screening tool in the prevention of sports-related injuries. However, the practical application of clinical foot posture evaluation indices in field settings remains limited due to the need for specialized knowledge and training among medical professionals. This study aimed to develop and validate a deep learning-based model that classifies foot posture using images of the hindfoot. The classification was based on indicators derived from the Foot Posture Index, focusing on the relative alignment of Abduction/Adduction of the forefoot on the rearfoot, as well

as Frontal plane alignment and position of the calcaneus. The model was designed to categorize foot posture into five types: adduction, abduction, varus, valgus, and neutral. A total of 750 posterior foot images were collected and divided into training (70%), validation (20%), and test (10%) datasets. During training, the model demonstrated progressive improvement, with increasing precision and decreasing loss values across epochs, indicating effective learning. The model achieved a mean Average Precision (mAP (B)) of 0.72 at IoU thresholds ranging from 0.50 to 0.95 (mAP50-95 (B)), and 0.80 at an IoU threshold of 0.50 (mAP50 (B)), reflecting high overall accuracy.

Evaluation using the test dataset revealed that while the classification accuracy for the neutral foot posture was relatively low (accuracy: 0.26), the model performed well in identifying the other four postures, with accuracy ranging from 0.69 to 1.00. These results suggest that the model has practical potential for use in sports and clinical settings, particularly for non-specialists who require accessible and automated tools for foot posture assessment. Future work will focus on improving the classification of neutral posture and expanding the dataset to enhance generalizability.

#### キーワード

足部形態, 深層学習, 判別モデル, 前足部内・外転, 踵骨アライメント

#### Keyword

Foot posture, Deep learning, Discriminant model, Forefoot adduction/abduction, Calcaneal Alignment

#### 要 旨

足部形態評価はスポーツ障害予防に有用であるが、医療従事者向けの足部形態評価の指標を現場で運用するには課題がある。

本研究ではDeep Learningを用い、足部形態評価の指標であるFoot posture indexの後方写真から判別できる指標を参考に内転・外転・内反・外反・通常の5つの足部形態を識別するモデルを開発した。750枚の画像を学習(70%)・検証(20%)・テスト(10%)に分割して開発と検証を進めた。適合率はエポック数の増加に伴い上昇し、損失関数は同様に減少する傾向を示し、適切にモデルが学習したことが確認され、mAP50-95=0.72,

mAP50=0.82と高精度を示した。

通常の足部画像の分類は正答率:0.26と低くなり精度の向上の必要性が課題にあがったが、4つの足部形態(内転, 外転, 内反, 外反)は、正答率が0.69~1.00と実用可能な精度であることが確認された。

#### 緒 言

スポーツ関連傷害の約20%は足部・足関節に関連する傷害であり<sup>1)</sup>、傷害予防のために様々な視点から対策が考えられている。対策のひとつとして、足部形態の評価があり、静止立位時の足部形態の評価は臨床的にも簡便に実施できることから利用されている。特に、過度な回内として知ら

れる扁平足は、後脛骨筋炎や脛骨過労性滑膜炎といったランニング障害との関連も報告されており<sup>2)</sup>、これらの足部形態の変化を早期に確認できる意義は高い。このような足部形態を評価できる指標のなかにFoot Posture Index (FPI)がある<sup>3)</sup>。この指標は、5つの視診と1つの触診から、-12～12までのスコアに分類することで足部形態を簡便に評価することが可能であり、特別な道具が必要でないことから、臨床現場やスポーツ現場でも導入しやすい。しかし、この指標は普段から足部・足関節の評価を実施する医療従事者にとっては簡便に評価可能であるが、実際にスポーツ現場で活動をする選手や監督にとっては、現場での活用の敷居が高いという課題も存在している。この課題解決に繋がる取り組みとして、5つの視診で可能な評価を写真で代替する方法があげられる<sup>4)</sup>。これにより、評価時間の短縮や現場にいない医療従事者が評価者として関わるができる可能性が示された。しかし、撮影された画像による評価を用いても評価者間での評価精度が十分でないという課題が残っている<sup>4)</sup>。

近年、Deep Learning技術を用いることで、適切な処理パラメータを決めることができるようになってきている。特に画像処理においては、畳み込みニューラルネットワーク (Convolutional Neural Network ; CNN) を活用することが多く、2012年の画像認識のコンテストILSVRCで優秀な結果を残している<sup>5)</sup>。この手法では、画像から得られる色、形状、テクスチャといった情報を用いて機械学習アルゴリズムを介して識別モデルを構築するため、評価者に依存せずに物体の認識と識別を高い精度で実施できると考えられている<sup>6)</sup>。実際、医療業界における医用画像分野では、医師の診断補助ツールとして導入される場面も増えており、肺結核の分類や腫瘍のサブタイプ分類など、医療資源の節約や医療の発展に貢献している<sup>7-9)</sup>。筆者らは、このような背景から、CNNの手法を足

部形態の評価に応用することで、これまでFPIを用いた足部形態評価で課題とされている検者間の評価精度や現場での活用の難しさを解決できる可能性があると考えた。しかし、現在までにこのような識別モデルは存在せず、検証に向けてモデルの開発と精度検証が必要となる。

本研究の目的は、1. FPI-6の後方写真から判断可能な「後足部に対する前足部の内外転」と「踵骨の前額面上の位置」を参考に前足部の内転・外転、踵骨の内反・外反、通常の5つの識別が可能なモデルを開発すること、2. 足部写真を用いた精度検証を行い、実用可能性を確認することとした。

## 1. 実験方法

### 1. 1 YOLOv11を用いた足部形態判別モデルのアルゴリズムの概要

本研究では、Ultralytics HUBチームによって開発されている物体検出アルゴリズム、You Only Look Once version 11 (YOLOv11)<sup>10)</sup>を用いて判別モデルを開発した。YOLOは、画像内の複数物体を同時に検出し、その位置とクラスを一括で出力できるアルゴリズムである。特にYOLOv11は、従来の検出方式と比較して、より軽量かつ高精度な物体検出を実現する最新のモデルである。画像識別の手法として広く用いられているCNNは、通常、画像全体をひとつのクラスに分類する。CNNの中でも代表的なモデル (Alex Net) を図1に示す<sup>5)</sup>。一方、YOLOは画像を格子状に分割し、区画ごとに物体の有無を判定し、存在する場合はその位置 (バウンディングボックス) とクラスを同時に推定する。YOLOの物体検知の仕組みを図2に示す<sup>11)</sup>。このようにして、画像中に複数の物体が存在する場合でも、それらの位置とラベルを一度に出力することが可能である。YOLOは、画像全体に対して一度のCNN推論でこれらの処理を完結できるため、従来手法に比べて高速

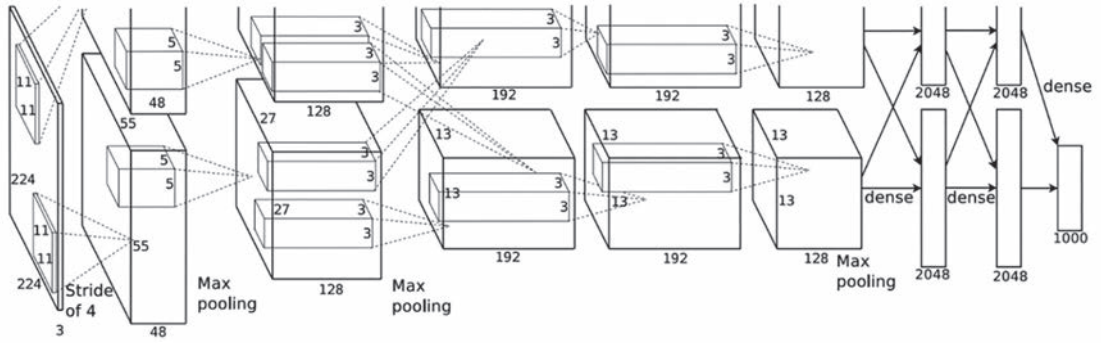


図1 CNNで代表的なAlexNetの構造<sup>5)</sup>

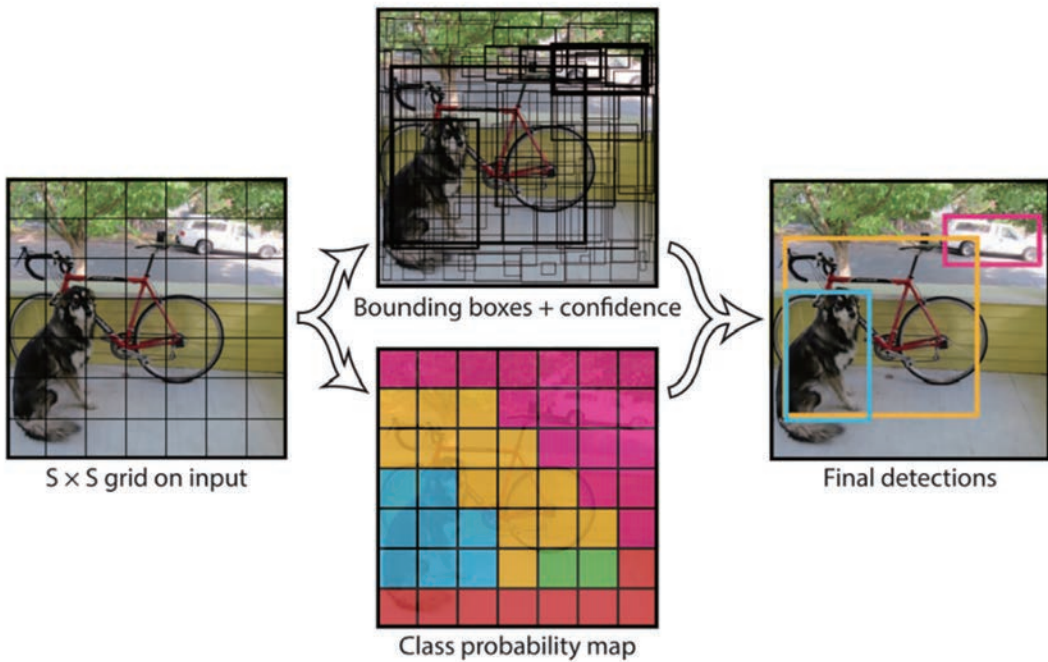


図2 YOLOによる物体検出の仕組み<sup>11)</sup>

かつ低コストで認識処理が可能である。このようなYOLOの特徴は、本研究で検証を実施するFPIという評価指標において、足部の形態を画像中の足首から下の領域のみで簡便に評価できる点で非常に有効であると考えた。物体検出アルゴリズムであるYOLOを用いることで、画像中から足部のみを自動で検出し、その領域に対して状態の識別（内転、外転、内反、外反、通常）を行うことが可能になる。また、画像処理においては、評価

対象となる領域が正確に選ばれていない場合、背景や衣類などのノイズの影響により、誤った判断がなされる危険性がある。特に、実験環境では問題がなくても、実際の運用環境においては、この影響が無視できない場合がある。したがって、評価対象の領域が適切に選択されているかをモデルが確認可能であることは、実運用を見据えた精度保証の観点から重要である。YOLOは、検出と分類を同時に行うことができるため、対象領域が適

切に選ばれたかどうかと、足の状態分類の結果を一度に得ることができる。このような構成により、正確性と効率性を両立した実験および運用が可能となる。

## 1. 2 データセットの作成

### 1. 2. 1 データサンプルの収集

対象は健康成人男性75名150足（平均±標準偏差：身長 $170.8 \pm 6.1$  cm，体重 $65.7 \pm 12.7$  kg，足長 $25.0 \pm 1.1$  cm，足幅 $9.8 \pm 0.5$  cm）とした。本研究は、ヘルシンキ宣言に基づき、所属機関の倫理委員会の承認を得て実施された（承認番号：19544-250424）。

対象ごとに「踵の位置は変えずに親指が見えるような足部位置（内転と定義）」、「踵の位置は変えず小指が見えるような足部位置（外転と定義）」、「踵をできるだけ内側に倒す姿勢（外反と定義）」、「踵をできるだけ外側に倒す姿勢（内反と定義）」、「通常の立位」の5条件で撮影を実施した。本研究に必要なデータサンプルの撮影は図3に示す撮影環境で実施した。撮影には、4台のアクションカメラ（GoPro12, GoPro Inc.）を用い、足部形態の撮影はすべてのカメラを同期して実施した。



図3 データサンプルの撮影環境

また、対象ごとに必要な足部・足関節の画像が取得できるようにカメラの設置距離と角度を設定した。具体的には、前方から撮影するカメラは高さ30 cmの三脚に固定したカメラを30°前方に傾けて設置し、内側から撮影するカメラは、足長の約50%の位置にカメラレンズの中心をあわせ、足部内側から踵骨-母趾がすべて写る位置に設置した。後方のカメラは踵骨の中心とカメラレンズの中心が一致する位置に設置し、後内側のカメラは踵骨に対して約30°の角度をつけて距舟関節が写る位置で撮影した。これらの撮影で得られた合計2100のデータサンプルのうち、本研究では後方からの撮影で得られた750のデータサンプルを使用して、「後足部に対する前足部の内転と外転」と「踵骨の前額面上の位置（内反と外反）」を通常とあわせて5つに識別するモデル構築に取り組んだ。なお、全750データの内訳は、5条件各150足であった。

### 1. 2. 2 データセットの準備、モデルの学習環境

本研究における足部形態識別モデルの学習は、Google Colabを用いて実施した。まず、収集した足部画像に対して、Roboflowを用いてアノテーションおよびデータセットの作成を行った。Roboflowでは、対象領域（足部）に対してバウンディングボックスを手動で設定し、各画像に「内転」、「外転」、「内反」、「外反」、「通常」のラベルを割り当てた。作成したデータセットは、YOLOv11形式でエクスポートした。また、学習・検証・テストの分割比（本研究では、学習画像70%、検証画像20%、テスト画像10%）や、画像の前処理（リサイズ）についても、Roboflow環境内で一貫して実施した。

次に、作成したデータセットをUltralytics社が提供するUltralytics HUBにアップロードし、クラス構成の視覚的確認、ラベル数の偏りの確認、学習用構成ファイル（.yaml）の自動生成を行っ

た。Ultralytics HUBはGUIベースの操作が可能であり、データセットの構造や設定内容を直感的に把握・管理できる点で有用である。また、エポック数、バッチサイズ、学習率などの学習条件を試行的に調整しやすい設計となっており、モデル構築前の準備段階を円滑に進めることができた。

モデルの学習は、Google Colab上のPyTorch環境にて実行した。Ultralyticsが提供するYOLOv11の公式実装を用い、エポック数100、バッチサイズ16、学習率0.01という条件下でトレーニングを行った。最適化手法にはAuto（自動選択）を採用し、モデルの性能向上を図った。学習の過程では、検証画像を用いた際の適合率（Precision）および損失関数（Loss）、再現率（Recall）、Intersection over Union Intersection over Union（IoU）閾値を条件設定した際の適合率といった主要な評価指標をエポックごとに出力し、学習の進捗を定量的に評価した。特に重要な指標の説明として、適合率はモデルが検出した対象のうち、正しく検出できた割合を示す指標であり、誤検出の少なさを反映する。一方、損失関数は、予測と正解ラベルとの誤差を示す指標であり、学習が進むにつれて減少することが望ましい。これらの指標がそれぞれ望ましい傾向を示している場合、モデルが適切に学習

されていると判断できる<sup>12)</sup>。

最終的に、学習中に一度も使用していないテスト用データを用いて、モデルの汎化性能を検証した。これにより、訓練されたモデルが未知の画像に対しても正確に足部形態を識別できるかを評価した。以上のように、本研究では、Roboflowによるデータセット構築、Ultralytics HUBによる設定支援、Google Colabによる学習実行と評価という役割分担のもと、クラウドツールを効果的に連携させることで、実用的かつ再現性の高いDeep Learning環境を構築した。ツール連携の流れを図4に示す。

## 結果

図5に、Google Colab上で100エポックの学習を行った際の適合率および損失関数の推移を示す。適合率はエポック数の増加に伴い上昇し、損失関数は同様に減少する傾向を示した。これにより、モデルが学習を通じて徐々に性能を向上させていることが確認された。Ultralyticsの評価指標に基づく詳細なメトリクスは以下の通りである：

metrics/precision (B) = 0.78, metrics/recall (B) = 1.00, metrics/mAP50-95 (B) = 0.72, metrics/mAP50 (B) = 0.82.

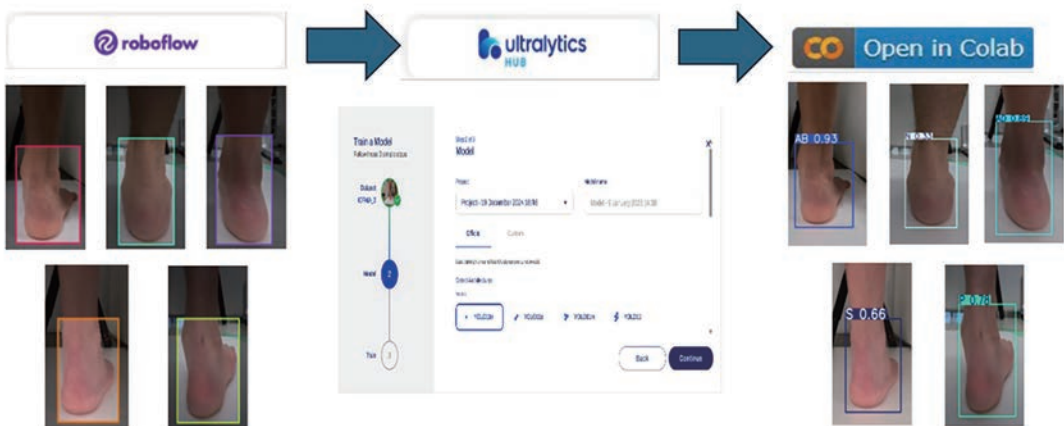


図4 モデル作成から学習までの流れ

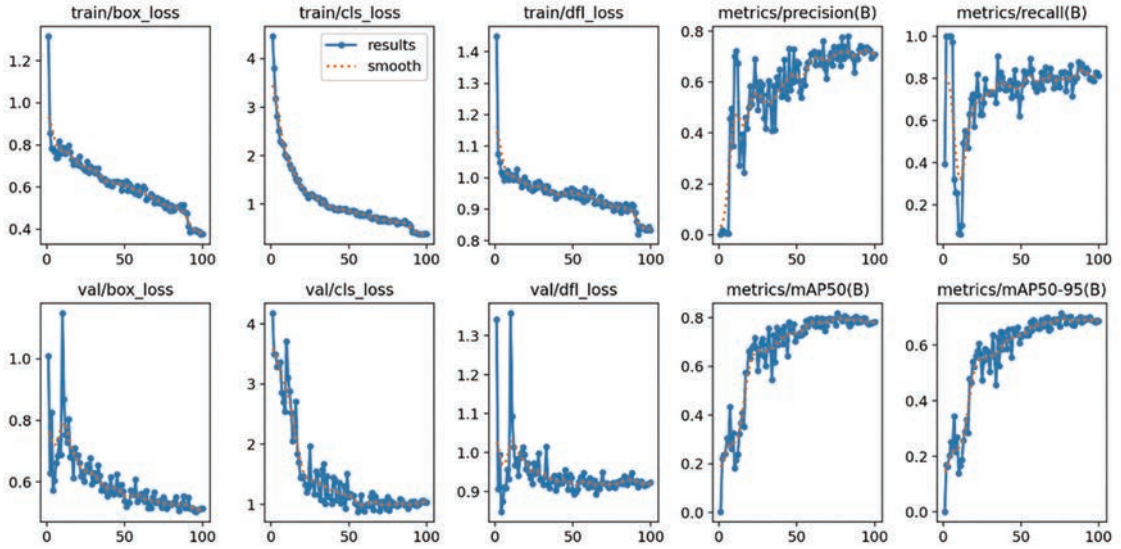


図5 Google Colabによる適合率 (Precision), 損失関数 (Loss) 率の変位

表1 テスト用データを用いた検証結果

	正解数	誤認識数	正解率
内転	15	0	1.00
外転	11	4	0.73
内反	11	3	0.85
外反	11	5	0.69
通常	4	11	0.26
平均	0.69	0.31	—

モデルの学習後、テスト用データを用いてモデルの汎化性能を評価した結果を表1に示す。全体の適合率は0.69、損失関数は0.31であった。誤認識の内訳としては、外転に属する4枚の画像が内転と誤認識された。通常に分類されるべき画像のうち、9枚が内反、2枚が外反と誤認識された。内反に分類されるべき画像のうち3枚が外反と誤認識され、外反に分類されるべき画像では、4枚が通常、1枚が内反と誤認識された。

### 考察

本研究では、YOLOv11を用いた足部形態識別モデルの構築と評価を行い、学習過程およびテストデータによる性能検証を通じて、モデルの有効性を確認した。本研究で得られた学習過程において、適合率がエポック数の増加に伴い上昇し、損

失関数は減少する傾向を示した。この結果は、モデルが学習を通じて徐々に識別性能を向上させていることを示しており<sup>12)</sup>、学習が安定して進行したことを裏付けるものである。

テスト用データによる評価結果から、全体の適合率が0.69、損失関数が0.31と算出された。さらに、Ultralyticsの評価指標に基づく metrics/precision (B) = 0.78, metrics/recall (B) = 1.00, metrics/mAP50-95 (B) = 0.72, metrics/mAP50 (B) = 0.82 という結果は、モデルが高い再現率を維持しつつ、精度および平均適合率においても良好な性能を示していることを意味する。特に、mAP50-95は、IoU閾値を0.50から0.95まで段階的に評価する厳密な指標であり、COCOやPASCAL VOCなどのベンチマークでも標準的に用いられている<sup>13,14)</sup>。mAP50-95 = 0.72は、最新の物体検出モデルの報告と比較しても同程度の数値を示しており<sup>15)</sup>、本研究のモデルが実用的な識別能力を有したものであると評価できる。

さらに、mAP50 (B) = 0.82 という結果は、バウンディングボックスの位置精度が高く、足部形態の空間的特徴を正確に捉えていることを示

している。Liらは、分類と位置推定の性能を分離して評価することの重要性を指摘しており<sup>16)</sup>、mAP50のような位置推定における信頼性の高い指標をとり入れることの重要性が強調されている。また、mAP50は視覚的に類似した対象の識別精度において検証すべき指標であることが示されている<sup>17)</sup>。これらの知見を踏まえると、mAP50 (B) で得られた値は、足部のように微細な形態差を識別するタスクにおいて、モデルの有効性を裏付ける精度の高さを示していると考えられる。

一方、誤認識の分析からは、特定のクラス間で混同が生じている傾向が確認された。外転では4枚の画像が内転と誤認識され、通常では9枚が内反、2枚が外反と誤認識された。内反では3枚が外反と誤認識され、外反では4枚が通常、1枚が内反と誤認識された。外転で内転と誤認識された画像では、小趾は出ているものの出ている面積が小さいことが視覚的に確認された。つまり、内外転の認識において、人が判断する際には母趾が見えているか、小趾が見えているかという基準で判断しているが、今回のモデルでは単に足の横幅の面積を判断基準にしている可能性があり、今後モデルを改良していくうえで検討課題になると考える。また、外反と内反を判断する際には、視覚的ではあるがアキレス腱の角度がバウンディングボックスからみて垂直に近い場合には通常、内側では内反、外側では外反と識別している傾向がみてとれたため、この点に焦点をあててアノテーションの仕方や撮影環境の調整を行うことで、精度向上につながる可能性があると考ええる。特に、通常の正解率が0.26と著しく低く、内反・外反の誤認が多く存在したためモデルの改善が必要となる。今回の研究では、通常の学習画像として通常の立位姿勢を設定したことが影響したと考えられる。通常の立位姿勢は、FPI-6のような足部形態評価での姿勢として用いられるなど、対象者の足

部の特徴を確認することができる点で有益であるが足部形態の通常とは定義が異なる。本研究では、そのような背景を踏まえたうえで、内反と外反という極端な設定を用いることで、通常を含めて内反と外反、通常を判別することができると考えてモデル開発に着手した。しかし、もともと外反の対象や内反の対象者が極端な条件で足部形態を変化させた場合では、判別閾値に影響するほどの足部形態の変化は生じず、これにより誤認識する確率が高まったのではないかと推察している。したがって、今後のモデル修正においては、踵骨が正常な条件での撮影を実施することやFPI-6の同じ分類項目内で0と判断される足部画像のみを学習に用いるなど、通常クラスの定義を再検討して修正と検証を進める必要があると考えている。

本研究にはいくつかの限界がある。はじめに、FPI-6では1つの項目ごとに-2～+2まで5つに分類してスコアをつける必要があるが、本研究では画像に対して、内転、外転、内反、外反、通常を認識して判別するモデルの開発にとどまっている。しかし、判別モデルを構築できたという点では新しい価値を提供しており、今後それぞれの特徴を学習させることで5つの分類モデルの作成とこれらの判別の組み合わせからスコア化を行いセルフチェックシステムの一般化に繋げることは可能と考える。また、対象が男性に限られているという点があげられる。女性の足部を対象としたサンプル画像の収集や、足のサイズが異なる対象など幅広い層をサンプルに組み込むことで、足部形態の評価精度を高めることにつながると思う。

#### まとめ

YOLOv11を用いて開発した足部形態識別モデルは、4つの足部状態（内転、外転、内反、外反）を実用可能な精度で識別可能であることが、検証画像とテスト画像の結果より確認できた。

## 謝 辞

本研究の遂行にあたり、研究助成を賜りました公益財団法人石本記念デサントスポーツ科学振興財団に厚く御礼申し上げます。

## 文 献

- 1) Lievers W.B., Goggins K.A., Adamic P., Epidemiology of foot injuries using national collegiate athletic association data from the 2009–2010 through 2014–2015 seasons., *J. Athl. Train.*, 55 (2) : 181-187 (2020)
- 2) Pérez-Morcillo A., Gómez-Bernal A., Gil-Guillen V. F., et al., Association between the Foot Posture Index and running related injuries: A case-control study., *Clinical Biomechanics*, 61: 217-221 (2019)
- 3) Redmond A.C., Crosbie J., Ouvrier R.A., Development and validation of a novel rating system for scoring standing foot posture: the Foot Posture Index., *Clinical Biomechanics*, 21 (1) : 89-98 (2006)
- 4) Terada M., Wittwer A.M., Gribble P.A., Intra- rater and inter- rater reliability of the five image- based criteria of the foot posture index- 6., *Int. J. Sports Phys. Ther.*, 9 (2) , 187-194 (2014)
- 5) Krizhevsky A., Sutskever I., Hinton G.E., Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 25, 1097–1105 (2012)
- 6) Zhao X., Wang L., Zhang Y., et al., A review of convolutional neural networks in computer vision., *Artif. Intell. Rev.*, 57 (4) , 99 (2024)
- 7) Paez R., Kammer M.N., Balar A., et al., Longitudinal lung cancer prediction convolutional neural network model improves the classification of indeterminate pulmonary nodules., *Scientific reports*, 13 (1) , 6157 (2023)
- 8) Jia H., Zhang J., Ma K., et al., Application of convolutional neural networks in medical images: a bibliometric analysis., *Quant. Imaging Med. Surg.*, 14 (5) , 3501-3518 (2024)
- 9) Yuan M., Ding H., Guo B., et al., Image-based subtype classification for glioblastoma using deep learning: prognostic significance and biologic relevance., *JCO Clin. Cancer Info.*, 8, e2300154 (2024)
- 10) Khanam R., Hussain M., Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv*, 2410, 17725 (2024)
- 11) Redmon J., Divvala S., Girshick R., Farhadi A., You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788 (2016)
- 12) Wang Q., Ma Y., Zhao K., Tian Y., A comprehensive survey of loss functions in machine learning., *Annals of Data Science*, 9 (2) , 187–212 (2022)
- 13) Wang W., Meng Y., Li H., Li S., Zhang C., Zhang G., Lei W., PDNet by Partial Deep Convolution: A Better Lightweight Detector., *Electronics*, 14 (3) , 591 (2025) .
- 14) Reis D., Kupec J., Hong J., Daoudi A., Real-time flying object detection with YOLOv8., *arXiv preprint, arXiv* : 2305.09972 (2023)
- 15) Wang Q., Wang Q.L., BT-YOLO11: Automatic Driving Road Target Detection in Complex Scenarios., *IEEE Access*, 13, 3562747 (2025)
- 16) Li Y., Zhu E., Chen H., Tan J., Shen L., Dense crosstalk feature aggregation for classification and localization in object detection., *IEEE Trans. Circuits Syst. Video Technol.*, 33 (6) , 2683–2695 (2022)
- 17) Li S., Li Y., Li Y., Li M., Xu X., YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection., *IEEE Access*, 9, 141861–141875 (2021)