

大規模言語モデルを活用した 経管栄養における看護行動認識の高精度化

Zhao Lingfeng¹ Garcia Christina¹ 小水内 俊介² コリー 紀代³ 佐藤 敦子⁴ 河内山 真由美⁴
那須 敏子⁴ 井上 創造¹

概要: 本研究の目的は、経管栄養（Gastrostomy Tube Feeding, GTF）の看護行動認識の精度を向上させることである。GTF における行動は、短時間の動作が長時間の動作に埋もれやすいことや、厳密な時系列制約が考慮されないといった課題がある。その結果、既存の手法では適切な認識が困難であり、より効果的な認識手法が求められている。本研究では、ビデオに基づくポーズ推定と大規模言語モデル（Large Language Model, LLM）を活用した新たな特徴量選択および後処理手法を提案する。具体的には、LLM を用いて時系列的な文脈を考慮した特徴量を自動生成し、これを手作業で設計した特徴量と統合することで、認識精度の向上を図る。さらに、多数決に基づく時間窓平滑化と短時間動作の優先処理を組み合わせた後処理手法を導入し、誤検出の低減を試みる。実験では、中国労災病院で収集した GTF データセットを用いて提案手法の有効性を検証した。このデータセットには、看護師、看護教員、看護学生の 3 グループが 3 日間にわたり実施した計 17 種類の動作が含まれている。実験の結果、従来の手法と比較して、F1 スコアが 53% から 66% へと向上し、後処理手法の導入により、最終的に F1 スコアは 68% に向上した。本研究は、看護師が行う多数のスキルの中でも特に誤認識リスクの高い経管栄養に注目し、その認識精度を向上させることで、教育現場における定量的かつ客観的な評価支援を実現する可能性を示した。今後は、より大規模なデータセットでの検証を進めるとともに、LLM を活用した動作認識のさらなる最適化や、リアルタイムでの動作識別の実現を目指す。

Improving Nursing Activity Recognition in Gastrostomy Tube Feeding Using Large Language Models

Lingfeng Zhao¹ Christina Garcia¹ Shunsuke Komizunai² Noriyo Colley³ Atsuko Sato⁴
Mayumi Kouchiyama⁴ Toshiko Nasu⁴ Sozo Inoue¹

1. はじめに

近年、医療・介護現場における看護師の行動認識に関する研究は、医療の質の向上や教育的フィードバックの自動化に向けて注目を集めている。中でも、経管栄養（Gastrostomy

Tube Feeding: GTF）に関する処置は、誤操作が患者の安全に直結するため、高度な技術と迅速な判断が求められる看護スキルの一つである。

しかし、経管栄養に関する動作は、一般的なバイタル測定や移乗支援などの基本的な看護動作と比べてデータセット上の登場頻度が少なく、既存の行動認識モデルでは十分な識別精度が得られていない。とりわけ、経管栄養における操作は短時間で動作が遷移し、個々の行動の粒度も高いため、精度の高い認識が困難である。

本研究では、経管栄養を対象とした行動認識の課題解決を目指し、以下の 3 点を重視している。

(1) 経管栄養は、医療安全上の配慮が特に求められる高度

¹ 九州工業大学
Kyushu Institute of Technology, Kitakyushu, Fukuoka 808-0196, Japan
² 香川大学
Kagawa University, Takamatsu, Kagawa 761-0396, Japan
³ 北海道大学
Hokkaido University, Sapporo, Hokkaido 060-0808, Japan
⁴ 広島文化学園大学
Hiroshima Bunka Gakuen University, Kure, Hiroshima 737-0004, Japan

な看護行為であること。

(2) 実施中の動作は短時間で次々と変化し、各動作の判別が難しいこと。

(3) 医療教育現場において、経管栄養シミュレータの教育評価指標の客観化・自動化が求められていること。

そこで本研究では、経管栄養に特化した教育用シミュレータ「Endotracheal Suctioning Training Environment Simulator - Tube Feeding (ESTE-TF)」[28]を用い、看護師、看護教員、看護学生の操作映像を記録し、行動認識モデルを構築した。映像から抽出した骨格データに、大規模言語モデル (LLM) による特徴設計を組み合わせ、認識精度の向上を図った。その結果、ベースラインと比較して F1 スコアは 54% から 66% へと大幅に改善された。

2. 関連研究

本章では、看護活動に関連する既存研究を概観し、本研究の動機となる課題を明確にする。著者らの先行研究 [27] に基づき、本稿ではデータセット拡充、時間窓条件追加、特徴量最適化を通じて、経管栄養における看護活動の認識精度向上を図った点が主な違いである。さらに、GTF および気管内吸引に関する研究動向、臨床における姿勢推定の課題、LLM を用いた特徴設計支援、後処理技術についても整理し、提案手法の意義を位置付ける。

2.1 看護行動認識の提案手法

看護師の行動認識は、医療の質の向上や技能評価の自動化を目的として、近年注目を集めている分野である。特に、気管内吸引や経管栄養といった医療処置に焦点を当てた研究が進められている。Islam らは、気管内吸引中の看護行動を認識するために、ビデオに基づく姿勢推定手法を提案した [1]。Ngo らは、多視点から得られた骨格データを用いて認識精度を向上させる手法や、技能評価との統合を試みた [2], [3]。さらに、姿勢推定と生成 AI の融合を目的とした看護行動認識チャレンジも開催されており [5]、その有効性が示されている。これらの研究は、姿勢推定の有用性を示す一方で、看護動作におけるデータの不均衡性や個人差の大きさといった課題も浮き彫りにしている。

2.2 姿勢推定と臨床環境の課題

臨床現場でのビデオに基づく姿勢推定は、遮蔽 (オクルージョン)、動きのブレ、カメラアングルの変動などの要因によって精度が低下する可能性がある [6], [8]。さらに、看護行動認識では、頻繁な姿勢の変化、複数人物の同時出現、および動作の短時間性がモデルの性能に大きく影響を与える。Dobhal ら [10] は、手の微細な動きや限られたデータが精度に与える影響を指摘している。そのため、精度の高い姿勢抽出に加え、動作の微妙な差異やラベルの曖昧さに対応可能な特徴量の設計が求められる。

2.3 特徴量の最適化手法

認識精度を高めるために、様々な特徴量設計が試みられている。基本的な骨格座標に加えて、関節角度、関節間距離、速度、加速度などの記述子が導入されてきた。気管内吸引の研究でも、加速度や動作軌跡を用いた評価が行われている [29], [30]。また、Kaneko および Inoue ら [12] は、大規模言語モデル (LLM) を用いて、センサ配置や特徴量の候補を自動提案し、従来の専門家依存の設計を軽減した。Ronando ら [13] や Shoumi ら [14] も、LLM を活用して疲労検出や行動認識の性能向上を図っており、従来法では見落とされがちな領域特化型特徴量を発見できる可能性を示している。

2.4 後処理技術

最適な特徴量を用いても、短時間動作や動作の境界付近では誤分類が発生しやすい。そのため、後処理は予測結果の滑らかさを保ち、誤認識を減少させる上で重要な役割を担う [17]。後処理の手法としては、多数決によるスムージングや、時間的局所化を調整するデータ駆動型の方法がある。Nag ら [15] は、行動検出における量子化誤差を補正する後処理手法を提案し、Tran ら [16] は未編集映像中の 3D 行動認識結果を改善する手法を示した。これらは、特に短時間動作や重複動作が多発する領域において有効性を発揮する。

2.5 LLM の応用と本研究の位置づけ

近年の LLM の進展により、医療支援システムにおいても文脈対応型の混合の手法が注目されている [19]。しかし、看護行動認識においては、姿勢抽出や個別の特徴量設計に重点を置いた研究が多く、LLM による特徴量提案と短時間動作への対応を同時に実現した先行研究はほとんど見られない。LLM におけるプロンプト設計は応答精度を左右する要素であり [18]、時間的文脈の適切な考慮は行動の識別・定位に不可欠である [21]。

本研究では、関節角度、速度、重心移動などの姿勢に基づく特徴量と、LLM によって得られた文脈的特徴量を融合し、さらに短時間動作に対応可能な適応型の後処理を組み合わせることで、医療環境における複雑で不均衡なデータへの頑健な対応を目指す。この提案手法により、看護行動認識の精度と解釈性の向上を図る。

3. データセット

本節では、経管栄養における看護活動のデータ収集と前処理プロセスについて詳述する。本研究では、最新の YOLO11 姿勢推定技術を用い、経管栄養施行中の看護活動に対し、2 次元関節点の抽出と解析を実施した。

さらに、参加者の人数、ビデオ記録の設定環境、ならびに操作中に注釈された具体的な動作カテゴリについて詳細

に説明する。

認識精度向上を目的に、骨格点の抽出結果に対して複数の前処理を施した。短時間動作の見落とし、人物重なりによる判別困難、ラベル分布の不均衡といった課題に対応するため、補間法による時系列データの平滑化・補完を行い、骨格データの連続性と信頼性を向上させた。

これらのデータ収集と処理は、特徴量抽出とモデル分析の基盤となり、短時間・高頻度行動認識における新たな提案手法を示すものである。

3.1 データ収集

本研究で使用したデータセットには、看護師3名、看護教員3名、看護学生3名を含む計9名の参加者による録画映像が含まれている。すべての参加者は3日間にわたり、計17種類の経管栄養に関する標準化された看護活動を実施した。各動作には対応するラベルが付与されており、その内容は表1に示している。なお、表1中の「その他」はGTFに関連しない操作を指す。

表1 経管栄養における看護活動

動作 ID	ラベル
0	患者への説明
1	必要物品の確認
2	手指の消毒
3	手袋の装着
4	栄養剤の準備
5	胃ろう部の状態確認
6	胃ろうキャップの開栓
7	微温湯の注入
8	栄養チューブの接続
9	注入速度の調整
10	手袋の取り外し
11	微温湯の準備
12	クレンメを閉じる
13	栄養チューブの取り外し
14	胃ろうキャップの開栓
15	使用済物品の片付け
16	その他

操作の標準化と実験結果の再現性を確保するために、本研究ではESTE-TF (Endotracheal Suctioning Training Environment Simulator – Tube Feeding) シミュレータを用いて看護活動を実施した。なお、模擬環境での実施であるため、頭部の挙上や薬剤投与といった一部のタスクは簡略化され、省略している。

録画映像は、1920 × 1080 ピクセル・33fps のデュアルカメラを搭載した UMIDIGI A3X スマートフォンにより撮影され、複数の視点から看護活動の一連の動作を正確に記録している。本研究の実施に先立って、倫理審査委員会の承認を得た (22-59)。参加者には研究目的と協力内容について口頭と書面にて説明し、書面による同意を得た。

3.2 短時間・類似動作の識別

経管栄養の中でも、短時間のクラスは特に難易度が高い。これらは手などの身体の一部に局在する最小限の動きで構成され、かつ短時間で発生するためである。さらに、これらの短時間の動作は、より長い時間枠で分析された場合に、しばしば他の動作に埋もれたり誤分類されたりする傾向がある。これは、長時間の動作から得られる運動パターンが時間的に集約されることで、信号が支配され、短時間動作に特有な微妙な遷移をモデルが識別しにくくなるためである。このような短時間動作の例は、図1に示されている。また、図1に示すように、(A) と (D) のように類似した動作が多数存在することも、認識タスクを一層困難にしている。



図1 短時間・類似動作の例

3.3 2次元姿勢推定

経管栄養の動作を分析するために、記録されたビデオから骨格情報を抽出するために YOLO11 モデル [9] を用いた。このモデルは、式 (1) に示すように、鼻、左目、右目、左耳、右耳、左肩、右肩、左肘、右肘、左手首、右手首、左腰、右腰、左膝、右膝、左足首、右足首の計17箇所の主要な身体の特徴点を識別し、それらを追跡することができる。YOLO11 による各フレームの姿勢推定出力は、各キーポイントの X 座標および Y 座標で構成されるベクトルとして表現される。

$$F = [x_0, y_0, x_1, y_1, \dots, x_{16}, y_{16}] \quad (1)$$

前処理の過程において、手術台によって遮られている膝や足首などの身体部位は除外された。これらの部位には、YOLO11 によってゼロ値が付与されており、映像内でこ

これらの関節点が検出されなかったことを示している。さらに、図2に示すようなデータの不均衡の問題に対処するため、本研究では合成少数過サンプリング技術（SMOTE）を用いてデータセットのバランスを調整した。SMOTEを用いることで、図に示されている頻繁かつ長時間の動作と、稀かつ短時間の動作の両方が公平に表現されるよう、データが拡張された。

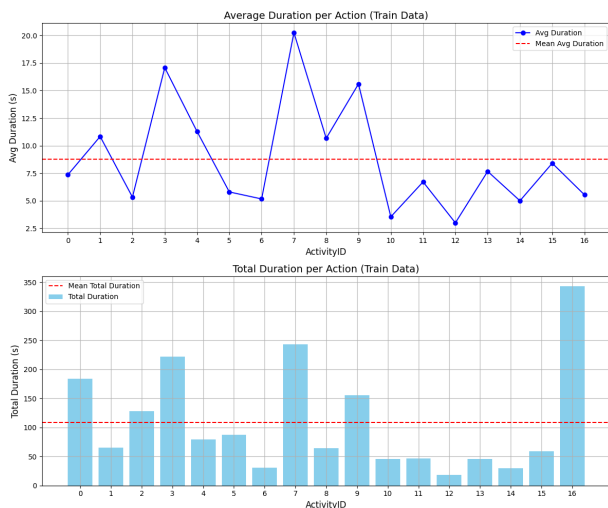


図2 学習データにおける各動作の平均時間および総時間

姿勢推定における重要な課題の一つは、1つのフレーム内に複数の人物が写っている場合である。例えば、図3に示す映像フレームでは、背景に静止した無関係の人物が写り込むことで、YOLO11によって誤検出が発生し、追加の骨格データが生成される可能性がある。この問題に対処するために、データ処理の段階で1620×1080ピクセルの事前定義された境界ボックス（バウンディングボックス）を適用し、検出領域を制限した。これにより、経管栄養中に看護師が実施するすべての動作を確実に捉えつつ、不要な検出を回避することが可能となった。

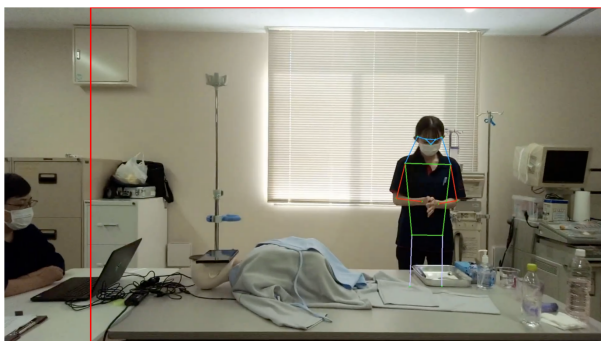


図3 複数人物が背景に含まれる例

3.4 補間法を用いたデータの平滑化

欠損した骨格点データの問題に対処するために、スムーズ長3秒の線形補間手法を用いた。この手法では、NumPy

のinterp関数を使用し、骨格点座標中のゼロ値を、周囲の非ゼロ値に基づく推定値で置き換えた。この手法により、欠損データが効果的に補完され、骨格点の軌跡における連続性が確保された。図4は、XおよびY座標の軌跡における平滑化の効果を示しており、急激な変化の低減とデータの一貫性の向上が確認できる。

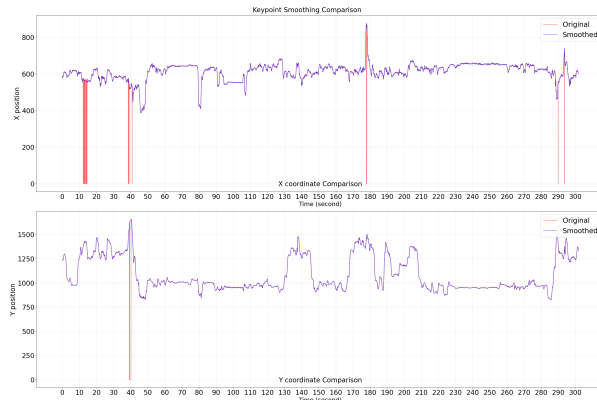


図4 補間によるデータ平滑化の可視化

4. 方法論

本節では、図5に示す手法について詳述する。本手法は、時系列および順序的文脈を考慮した特徴量を得るために大規模言語モデル（LLM）を活用し、さらにクラス順の不整合を補正するためのターゲット型後処理を組み合わせることで、経管栄養（GTF）における動作の精度を向上させることを目的としている。

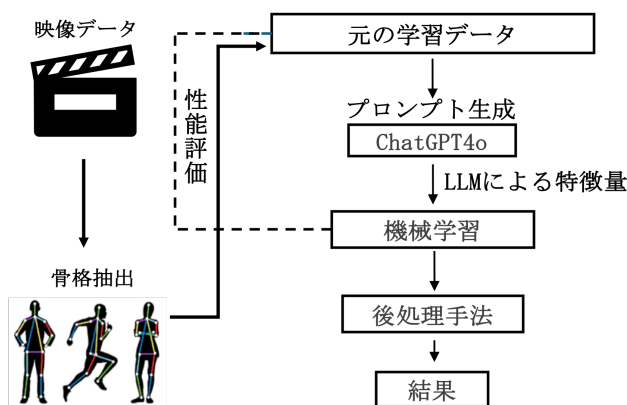


図5 経管栄養行動のパイプライン

4.1 LLM手法におけるプロンプト設計

経管栄養に関連する活動の認識精度を高めるためには、短時間および長時間のサブアクティビティの両方を正確に識別する必要がある。手作業による特徴量設計では、特に胃ろうキャップの開栓・閉栓のように、動きが最小限かつ手元に局在している場面では、微細ではあるが重要な運動の手がかりを見落とす可能性がある。さらに、動作の順序

の不整合や一部動作の見逃しは、医療分野における姿勢推定手法の有効性を制限する要因となっている。

そこで本研究では、学習データから得られる情報と文脈を活用したプロンプティングにより、LLM が提案する新たな特徴量を生成し、それらを手作業で設計した特徴量と組み合わせて、認識モデルの最適化を図る。Few-shot プロンプティングの出力は、式 (2) で数学的に表現される。活動に関連する主要なキーポイントを考慮することで、映像から抽出された姿勢情報の最適化を実現している。

$$F^* = \arg \max_F \mathbb{P}_{\text{LLM}}(F | P) \quad (2)$$

$$P = \text{Prompt}(R, C, E, T) \quad (3)$$

$$F = \{f_i \mid f_i \in G(P), i = 1, \dots, n\} \quad (4)$$

本研究では、式 (3) に示す P に含まれる時間的・系列的コンテキストを用いて、few-shot prompting を適用し、式 (4) に示される LLM 特徴量候補 F の最適化を行った。LLM モデルとしては、OpenAI によって提供される GPT-4o を用い、Web インターフェース経由でアクセスされた。その際、初期設定（温度 0.7、最大トークン制限あり、top-p = 0.9~1.0 の nucleus sampling）を採用した。この LLM を活用することで、強力なコンテキスト理解能力に基づき、文脈に依存した詳細かつ適切な特徴量提案が可能となる。

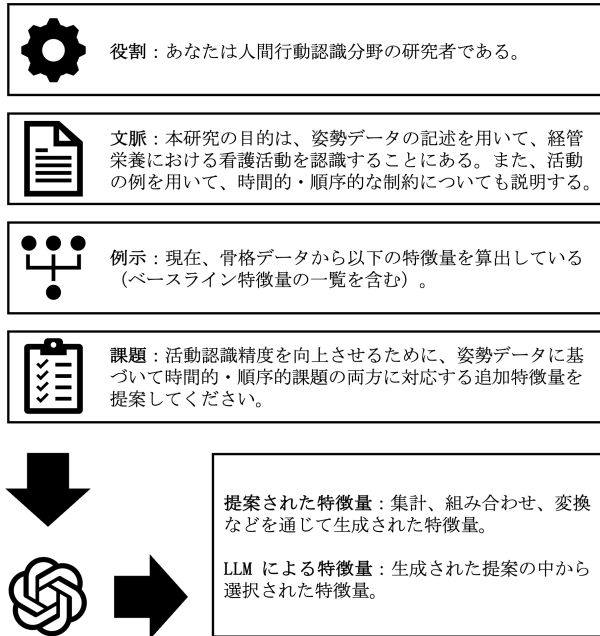


図 6 経管栄養活動特徴量に対する LLM プロンプト設計手法

図 6 に示すように、経管栄養活動中の骨格データから有効な特徴量を抽出するために用いられたプロンプト手法は、(1) 役割、(2) 文脈、(3) 例示、(4) 課題という4つの構成要素から成り立っており、それぞれが大規模言語モ

デル（LLM）による適切かつ文脈依存的な特徴量提案を導く上で重要な役割を果たしている。

- **役割** モデルに対してドメイン固有の視点を与え、出力される LLM の応答が関連性のあるものとなるようにする。
- **文脈** タスク、データ、および課題に関する詳細な背景情報を提供し、モデルが問題空間を正確に理解できるようにする。
- **例示** 重複や無関係な提案を防ぐために、既存の特徴量の基準やガイドラインを示す。
- **課題** モデルに期待される出力とその目的を明確に定義する。

4.2 基本特徴量による行動認識

経管栄養活動の骨格データに基づく分類には、過学習に対する耐性と汎化性能の高さから、ランダムフォレスト分類器（Random Forest Classifier : RFC）[20] を採用した。RFC は非線形データの処理に優れており、複雑な関係性を持つ特徴量に対しても柔軟に対応できるため、人間の行動認識データセットでよく見られる数値的およびカテゴリー的特徴量の両方を効果的に扱うことができる。

本研究では、RFC モデルの堅牢性をさらに高めるために、GridSearchCV を用いてハイパーパラメータの最適化を実施し、GTF データセットに最適な構成を探索した。最終的に選定された最良モデルのパラメータは以下の通りである。n_estimators は 500 に設定し、決定木を増やすことでアンサンブル全体の安定性を高めた。max_depth は None に設定し、各決定木が最大限に成長できるようにした。max_features は sqrt とし、分岐ごとに考慮する特徴量の数を特徴量数の平方根に制限した。min_samples_leaf は 1、min_samples_split は 2 に設定し、細粒度な分類を可能とするためにきめ細かい分割を促進した。さらに、bootstrap は False に設定し、すべての特徴量を置換なしで使用する構成とした。また、学習の高速化のために n_jobs を -1 に、結果の再現性を確保するために random_state を 42 に設定している。これらの最適化されたパラメータにより、RFC モデルは経管栄養活動の骨格データに含まれる複雑なパターンを高精度に捉えることが可能となった。

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (5)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (6)$$

$$\max_ft = \max_i x_i \quad (7)$$

$$\min_ft = \min_i x_i \quad (8)$$

$$var_ft = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (9)$$

$$med_ft = median(X) \quad (10)$$

$$sum_ft = \sum_{i=1}^N x_i \quad (11)$$

$$\cos(\theta) = \frac{\mathbf{BA} \cdot \mathbf{BC}}{|\mathbf{BA}| \cdot |\mathbf{BC}|}, \quad \theta = \arccos(\cos(\theta)) \quad (12)$$

$$v_i = x_{i+1} - x_i, \quad v_seg = \{v_1, v_2, \dots, v_{N-1}\} \quad (13)$$

経管栄養活動の骨格データから抽出された基本特徴量は、統計的、幾何学的、および動作に基づく記述子で構成される。統計的特徴量には、平均 (μ , 式 (5)), 標準偏差 (σ , 式 (6)), 最大値 (max_ft , 式 (7)), 最小値 (min_ft , 式 (8)), 分散 (var_ft , 式 (9)), 中央値 (med_ft , 式 (10)), および総和 (sum_ft , 式 (11)) が含まれ、データの中心傾向や分布、集計値を捉えることができる。

幾何学的特徴量には、身体部位間の相対的な向きを表現するために、ベクトル射影を用いて計算される関節角度 (θ , 式 (12)) が含まれる。さらに、動作に基づく特徴量として、連続フレーム間の位置差を表す速度 (v_i , 式 (13)) が導出されている。

これらの特徴量は、動作および姿勢の詳細な分析を可能にし、複雑な経管栄養における看護活動の認識において強力な基盤を提供する。学習用データセットには看護師・看護教員・看護学生各2名の計6名、テスト用には各属性から1名ずつ、計3名による映像データを使用された。

4.3 多数決に基づく時間窓平滑処理

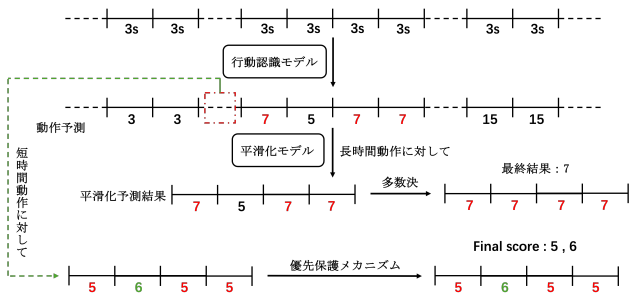


図7 経管栄養活動における後処理手法のフレームワーク

図7は、多数決に基づく時間窓平滑処理を特徴とする、経管栄養における看護活動データに対して提案された後処理手法の詳細を示している。継続時間に起因する予測の不整合を軽減するために、短時間動作 S の認識精度を維持しつつ、長時間動作 L によって生じるラベルの繰り返しを補正する目的で、優先的な処理を導入している (式 (14) および式 (15))。

すべての連続フレームが行動としてラベル付けされたセグメント $Seg(a)$ と、モデルによる予測ラベル y_{pred} を用いた後処理は、長時間動作に対しては式 (16), 短時間動作に対しては式 (17) により数学的に表現される。

$$L = \{a \in A \mid duration(a) > T\} \quad (14)$$

$$S = \{b \in A \mid duration(b) \leq T\} \quad (15)$$

$$y_{post}[i] = mv(y_{pred}[seg(a)]), \quad if \ y_{pred}[i] \in L \quad (16)$$

$$y_{post}[i] = y_{pred}[i], \quad if \ y_{pred}[i] \in S \quad (17)$$

5. 結果

本節では、大規模言語モデルにより生成された特徴量の貢献、後処理が認識精度に与える影響、さらに短時間および長時間動作の認識性能の向上について分析する。これらの結果評価は、Baseline モデルとの比較指標を用いて実施する。

表2 異なるプロンプト文脈による特徴量

アプローチ	特徴量
Baseline	基礎統計特徴, 関節角度, 速度
Few-shot(FS1)	関節間距離, 腰部間距離, 肩部間距離, 加速度
Few-shot(FS2)	加加速度
Few-shot(FS3)	重心座標, 手の間距離, 頭部間距離

各プロンプトにより生成された特徴量は、関連性に基づいて選別され、前段階の特徴量セットに逐次的に加えられていく。最終的には、すべての有用な特徴量を包含する包括的な特徴量セットが構築される。この手法により、各プロンプトがもたらす固有の情報を保持しつつ、文脈に応じた生成を組み合わせることで、より網羅的かつ表現力の高い特徴量セットを形成する利点が際立つ。

5.1 特徴量を用いた行動認識

同じ特徴量の組み合わせに対して、異なる時間窓サイズとオーバーラップ率を適用し、経管栄養における行動認識性能への影響を詳細に評価した。特に、時間窓サイズの違いが分類性能に及ぼす影響を明確にすることを目的とし、時間スケールの違いに応じて行動認識精度がどのように変化するかを検証した。

最初に、表3は、WINDOW.SIZEを2秒、OVERLAP.RATEを0.5に設定した条件での評価結果を示す。この設定は時間分解能が高く、短時間で発生する動作や一時的な動作の検出に有効である。実験では、Baseline 特徴量と、各 Few-shot プロンプト (FS1~FS3) で拡張された特徴量セットによる分類性能を比較し、それぞれの特徴量

が認識精度に与える効果を定量的に分析した。この結果から、提案手法の有効性と、時間的設定との相互作用を確認することができた。

表 3 認識結果 (WINDOW_SIZE=2 秒, OVERLAP_RATE=0.5)

動作 ID	Base	Base*	FS1*	FS2*	FS3*
0	0.41	0.45	0.48	0.48	0.51
1	0.73	0.82	0.76	0.77	0.75
2	0.45	0.50	0.54	0.55	0.68
3	0.64	0.62	0.64	0.65	0.71
4	0.62	0.63	0.67	0.66	0.69
5	0.42	0.44	0.43	0.45	0.51
6	0.00	0.00	0.00	0.00	0.00
7	0.66	0.72	0.75	0.76	0.78
8	0.40	0.47	0.44	0.49	0.51
9	0.91	0.91	0.91	0.91	0.91
10	0.11	0.36	0.35	0.35	0.19
11	0.84	0.81	0.78	0.79	0.79
12	0.00	0.25	0.00	0.00	0.00
13	0.12	0.39	0.39	0.29	0.25
14	0.00	0.30	0.25	0.25	0.26
15	0.46	0.48	0.44	0.48	0.55
16	0.45	0.50	0.56	0.54	0.56
合計 F1 スコア	0.53	0.58	0.59	0.60	0.62

※「*」は、SMOTE を使ってデータを拡張した分類結果である。

続いて、表 4 は、WINDOW_SIZE を 3 秒、OVERLAP_RATE を 0.8 に設定した結果を示す。この条件は、より広い時間的文脈をモデルに提供し、特に長時間動作の識別精度向上を目的としている。異なるウィンドウ設定を通じて、提案手法における特徴量の有効性と頑健性を多面的に検証した。

表 4 認識結果 (WINDOW_SIZE=3 秒, OVERLAP_RATE=0.8)

動作 ID	Base	Base*	FS1*	FS2*	FS3*
0	0.42	0.49	0.54	0.55	0.57
1	0.78	0.79	0.80	0.80	0.80
2	0.67	0.54	0.71	0.66	0.75
3	0.69	0.68	0.71	0.69	0.73
4	0.58	0.65	0.67	0.61	0.63
5	0.52	0.52	0.59	0.44	0.59
6	0.00	0.00	0.00	0.00	0.00
7	0.70	0.74	0.75	0.73	0.79
8	0.41	0.47	0.44	0.42	0.42
9	0.93	0.92	0.92	0.92	0.90
10	0.00	0.38	0.15	0.15	0.40
11	0.81	0.87	0.89	0.89	0.87
12	0.00	0.00	0.00	0.00	0.00
13	0.10	0.16	0.13	0.20	0.20
14	0.00	0.32	0.28	0.28	0.28
15	0.52	0.47	0.53	0.58	0.64
16	0.53	0.48	0.58	0.55	0.63
合計 F1 スコア	0.59	0.61	0.64	0.62	0.66

※「*」は、SMOTE を使ってデータを拡張した分類結果である。

5.2 後処理

経管栄養における行動認識における系列的な予測誤りに対処するために、本研究では、多数決に基づく時間窓平滑化と短時間動作の優先処理を組み合わせた後処理手法を検討した。

多数決に基づく時間窓平滑化 (P1) は、予測ラベルのばらつきを抑える効果があるが、ID 6, 10, 12, 14 などの短時間動作が他の動作に埋もれてしまうリスクがある。この課題に対し、短時間動作の優先処理 (P2) では、短時間動作の ID およびその最小持続時間を事前に定義することで、これらの動作を保護しつつ、P1 による平滑効果を活用して長時間動作の識別精度も維持することを可能にする。

これらの処理を組み合わせることで、短時間および長時間の動作が混在する経管栄養プロセスにおいて、より安定かつ正確な動作認識を実現する。表 5 に、各後処理手法の F1 スコアを示す。

表 5 後処理手法による行動認識性能

動作 ID	Base	P1	P2
0	0.42	0.56	0.56
1	0.78	0.81	0.81
2	0.67	0.72	0.73
3	0.69	0.76	0.77
4	0.58	0.68	0.68
5	0.52	0.72	0.73
6	0.00	0.00	0.00
7	0.70	0.78	0.78
8	0.41	0.46	0.45
9	0.93	0.93	0.93
10	0.00	0.00	0.40
11	0.81	0.87	0.87
12	0.00	0.00	0.00
13	0.10	0.10	0.18
14	0.00	0.28	0.27
15	0.52	0.81	0.81
16	0.53	0.65	0.65
合計 F1 スコア	0.59	0.68	0.68

P2 は、短時間の動作（例えば“胃ろうキャップの閉栓” (F1 スコア = 0.27) および“手袋の取り外し” (F1 スコア = 0.40)) の認識精度を大幅に向上させた。これらの動作は、Baseline においては F1 スコアがいずれも 0.00 であり、モデルが十分に検出・識別できていなかったことが示されている。P2 の導入により、これらの短時間かつ局所的な動作に対する検出性能が大きく改善された。

全体的な結果として、P2 は多くの活動において最も高い F1 スコアを一貫して維持しており、最終的な総合 F1 スコアは 0.68 に達した。これは、Baseline で得られた F1 スコア 0.59 と比較して大きな性能向上を示しており、提案手法の有効性を裏付けている。

さらに、混同行列の比較によって、短時間活動に関する

誤分類の減少が明確に確認された。図 8 は、Baseline モデルが短時間の動作をうまく区別できず、誤認識している様子を示しているのに対し、図 9 は、P2 の後処理を適用することで認識精度が著しく向上し、特に短時間動作に対する分類性能が大幅に改善されたことを示している。

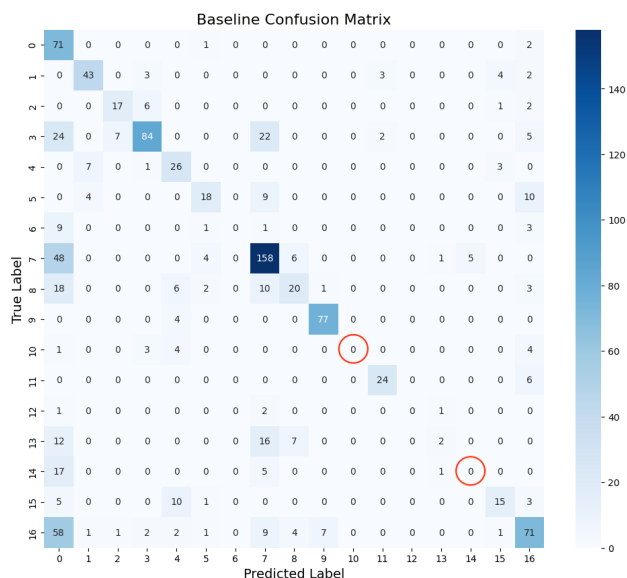


図 8 Baseline の混同行列

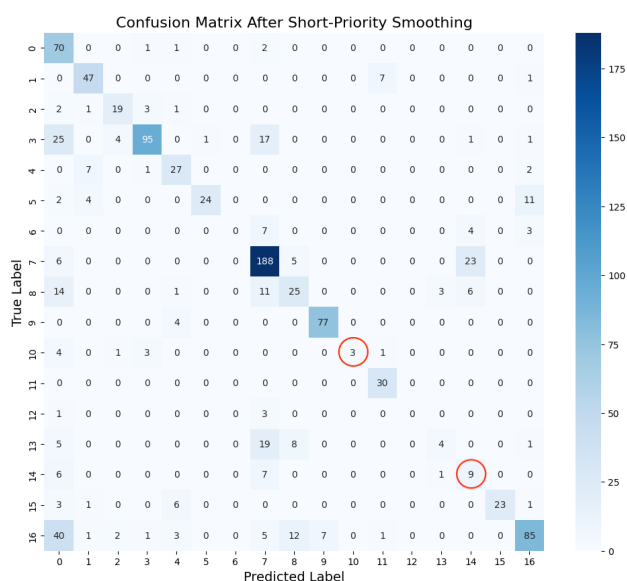


図 9 Few-Shot における P2 後処理の混同行列

6. 経管栄養活動に関する考察

6.1 経管栄養看護動作の特性と課題

経管栄養（Gastrostomy Tube Feeding, GTF）は、日常生活動作と比較して動作の順序性、動作時間のばらつき、部分的な遮蔽といった複雑な特徴を持つため、行動認識において特有の課題を伴う。さらに、「手指の消毒」「手袋の

装着」「栄養チューブの接続」など一連の手順で構成され、動作の順序性が非常に重要である。数秒程度の短時間動作（例：胃ろうキャップの開栓・閉栓）と、数十秒以上の長時間動作が混在しており、モデルには短時間動作の微細な変化と長時間動作の時間的文脈の両方を捉える能力が求められる。

多数決に基づく時間窓平滑化は、ノイズ除去や長時間動作の影響緩和に有効であるが、GTF に特有の時間構造には限界がある。具体的には、(1) 動作の重複や並列実行、(2) 階層的な動作依存、(3) 長時間動作中の瞬間的キーフレームなどに対して、局所的な時間範囲のみを対象とする手法では適切に対応することが困難である。

今後は、Transformer や Graph Neural Network (GNN) などの高度な時系列モデリング手法を導入し、動作の論理的関係や階層構造、長期的依存性を明示的にモデル化することで、より精緻な GTF 動作認識が期待される。

6.2 一般的な看護・非医療動作データとの比較

今回の動画撮影時の条件として、経管栄養シミュレータを用いたトレーニングであったため、厳密な順序性（技術の正確性）が求められ、臨床経験を有する看護師と学生とで所要時間に格差（動作時間の極端なバラツキ）があった。また、手の動作が器具などによって遮蔽されやすいという特徴もあった。これらは従来の動作認識モデルにとって大きな課題である。さらに、経管栄養では「胃ろうキャップの開閉」や「クレンメを閉める」など、瞬間的かつ微細な動作の認識が求められ、これは姿勢推定の精度や補間処理にも大きく依存する。

また、経管栄養の看護活動では 1 つの動作カテゴリが複数のマイクロラベルに分割されており、これが詳細な監督信号を提供する一方で、アノテーションおよび行動認識の複雑性を増加させている。これにより、より精緻な時間情報の処理や特徴量抽出、後処理最適化が要求される。

また、経管栄養とともに医療的ケアに分類される気管内吸引（Endotracheal Suctioning）と比較すると、本研究における経管栄養看護活動のデータ収集では、同時動作や遮蔽の影響がより顕著であり、瞬間的な動作の抽出が一層困難であった。

6.3 LLM を用いた特徴量生成の貢献

大規模言語モデル（LLM）による特徴量生成は、従来の人工的特徴量に比べて、経管栄養における動作の識別性能を大きく向上させた。特に、関節速度や角度の変化といった手元の動作に関する特徴量は、胃ろうキャップの開栓・閉栓など短時間の動作認識に効果的であった。大規模言語モデルは、自然言語によるプロンプトを通じて、動作の時間的スケールに応じた特徴量（例：短時間動作に対する加速度や角速度、長時間動作に対する順序依存性など）を自

動で提案できる。

また、LLM は事前学習によって獲得した医療領域の知識を活用し、データ不足や新規タスクにおいても汎化性能を発揮する。これにより、微細な動作の認識精度向上や、特徴空間の最適化が可能となった。

本研究における LLM 活用は、経管栄養における特有の動作構造に対応するための革新的な提案手法であり、従来手法では困難だった短時間動作の捉え方や時間整合性の保持に貢献した。

7. まとめ

本研究では、大規模言語モデル (LLM) によるプロンプト設計における文脈の違いが、看護行動認識における特徴量生成とモデル性能に与える影響を検証した。実験の結果、プロンプトの文脈が生成される特徴量に大きく影響し、その結果として認識精度にも差が生じることが明らかとなった。

さらに、ビデオに基づく姿勢推定、LLM を用いた特徴量抽出、および後処理手法を統合した包括的な経管栄養 (Gastrostomy Tube Feeding, GTF) 行動認識フレームワークを提案した。YOLO11 により骨格情報を抽出し、SMOTE によりクラス不均衡を補正した。LLM によって生成されたドメイン固有の特徴量を人工的特徴量と統合し、ランダムフォレスト分類器により認識を行った。さらに、短時間動作への優先処理を含む後処理により、予測精度の向上を図った。

さらに、経管栄養活動の認識モデルの構築にあたり、合計 9 本のビデオデータを使用した。トレーニングデータとしては、看護師 2 名、看護教員 2 名、看護学生 2 名による計 6 本の映像を用い、テストデータとしては、看護師、看護教員、看護学生の各 1 名ずつによる計 3 本の映像を使用した。その結果、提案手法は Baseline と比較して F1 スコアが 54% から 66% に改善された。また、後処理手法の導入により、最終的に F1 スコアは 68% に向上した。

今後は、Transformer や Graph Neural Network (GNN) などの高度な時系列モデルや、GAN などのデータ拡張技術の導入により、時間依存性の強い看護活動への対応力を強化する必要がある。また、より多様な医療手技への適用と、実環境における汎用性の検証も重要な課題である。

さらに、BioGPT や MedPaLM などのドメイン特化型 LLM を柔軟に統合できる手法を検討し、経管栄養における動作認識の精度をさらに向上させることが求められる。

参考文献

- [1] Islam, S., Hossain, S. M. H., Uddin, M. Z., Hossain, S., and Ahad, M. A. R. "Enhancing Nursing Activity Recognition During Endotracheal Suctioning Through Video-based Pose Estimation and Machine Learning". *International Journal of Activity and Behavior Computing*, 2024(3), pp. 1–15. <https://doi.org/10.60401/ijabc.36>
- [2] Ngo, H. A. V., Vu, Q. N. P., Colley, N., Ninomiya, S., Kanai, S., Komizunai, S., Konno, A., Nakamura, M., and Inoue, S. "Toward recognizing nursing activity in endotracheal suctioning using video-based pose estimation". *International Journal of Activity and Behavior Computing*, 2024(1). <https://doi.org/10.60401/ijabc.1>
- [3] Ngo, H. A. V., Colley, N., Ninomiya, S., Kanai, S., Komizunai, S., Konno, A., Nakamura, M., and Inoue, S. "Nurses' Skill Assessment in Endotracheal Suctioning Using Video-based Activity Recognition". *International Journal of Activity and Behavior Computing*, 2024(2), pp. 1–24. <https://doi.org/10.60401/ijabc.20>
- [4] Garcia, C., and Inoue, S. "Challenges and Opportunities of Activity Recognition in Clinical Pathways." In *Human Activity and Behavior Analysis*. 2024. CRC press. <https://doi.org/10.1201/9781003371540-8>
- [5] Ngo, H. A. V., Kaneko, H., Hassan, I., Ronando, E., Shoumi, M. N., Munemoto, R., Hossain, T., and Inoue, S. "Summary of the nurse care activity recognition challenge using skeleton data from video with generative AI". *International Journal of Activity and Behavior Computing*, 2024(3), pp. 1–20. <https://doi.org/10.60401/ijabc.31>
- [6] Liu, J., Mu, X., Liu, Z. et al. "Human skeleton behavior recognition model based on multi-object pose estimation with spatiotemporal semantics". *Machine Vision and Applications* 34, 44(2023). <https://doi.org/10.1007/s00138-023-01396-0>
- [7] Inoue, S., Lago, P., Hossain, T., Mairitha, T., and Mairitha, N. Integrating Activity Recognition and Nursing Care Records: The System, Deployment, and a Verification Study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 86 (September 2019), 24 pages. <https://doi.org/10.1145/3351244>
- [8] Morais, R., Le, V., Tran, T., Saha, B., Mansour, M., and Venkatesh, S. "Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 11988–11996. doi: 10.1109/CVPR.2019.01227.
- [9] Khanam, R. and Hussain, M. "YOLOv11: An Overview of the Key Architectural Enhancements". 2024. doi: 10.48550/arXiv.2410.17725.
- [10] Dobhal, U., Garcia, C., and Inoue, S. "Synthetic Skeleton Data Generation using Large Language Model for Nurse Activity Recognition". In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '24)*. Association for Computing Machinery, New York, NY, USA, pp. 493–499. <https://doi.org/10.1145/3675094.3678445>
- [11] Miyake, N., Kaneko, H., Ronando, E., Garcia, C., Inoue, S. "Toward Detecting and Explaining Stress of Nurses Using Wearable Devices and LLMs". *Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2024)*. UCAmI 2024. Lecture Notes in Networks and Systems, vol 1212. Springer, Cham. https://doi.org/10.1007/978-3-031-77571-0_28
- [12] Kaneko, H., and Inoue, S. "Toward Pioneering Sensors and Features Using Large Language Models in Human Activity Recognition". In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, pp. 475–479. <https://doi.org/10.1145/3594739.3610741>

- [13] Ronando, E., and Inoue, S. "Improving Fatigue Detection with Feature Engineering on Physical Activity Accelerometer Data Using Large Language Models". *International Journal of Activity and Behavior Computing*, 2024(2), pp. 1–22. <https://doi.org/10.60401/ijabc.18>
- [14] Shoumi, M. N., and Inoue, S. "Leveraging the Large Language Model for Activity Recognition: A Comprehensive Review". *International Journal of Activity and Behavior Computing*, 2024(2), pp. 1–27. <https://doi.org/10.60401/ijabc.21>
- [15] Nag, S., Zhu, X., Song, Y-Z., and Xiang, T. Post-processing temporal action detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 18837–18845, 2023. <https://doi.org/10.1109/CVPR52729.2023.01806>
- [16] Tran, M. T., Vu, M. Q., Hoang, N. D., and Bui, K-H. N. "An Effective Temporal Localization Method with Multi-View 3D Action Recognition for Untrimmed Naturalistic Driving Videos," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 2022, pp. 3167–3172, doi: 10.1109/CVPRW56347.2022.00357.
- [17] C. Wang, Y. Xu, H. Liang, W. Huang and L. Zhang, "WOODY: A Post-Process Method for Smartphone-Based Activity Recognition," in *IEEE Access*, vol. 6, pp. 49611–49625, 2018, doi: 10.1109/ACCESS.2018.2866872.
- [18] Amatriain, X. "Prompt Design and Engineering: Introduction and Advanced Methods." *arXiv preprint arXiv:2401.14423*, 2024. <https://arxiv.org/abs/2401.14423>
- [19] Javaheri, H., Ghamarnejad, O., Lukowicz, P., Stavrou, G. A., and Karolus, J., "LLMs Enable Context-Aware Augmented Reality in Surgical Navigation" , Art. no. arXiv:2412.16597, 2024. doi:10.48550/arXiv.2412.16597.
- [20] Gangwal, R. Types of Sampling and Sampling Techniques. *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2019/09/data-scientists-guide-8-types-of-sampling-techniques/>. Accessed July 10, 2024.
- [21] Alwassel, H., Heilbron, F. C., Escorcia, V., and Ghanem, B. Diagnosing Error in Temporal Action Detectors. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part III*. Springer-Verlag, Berlin, Heidelberg, 264–280. https://doi.org/10.1007/978-3-030-01219-9_16
- [22] Li, Z., Deldari, S., Chen, L., Xue, H., and Salim, F. D. SensorLLM: Aligning large language models with motion sensors for human activity recognition. *OpenReview*. 2025. <https://openreview.net/forum?id=cDd7kg9mkP>
- [23] Xi Chen, Julien Cumin, Fano Ramparany, Dominique Vaufreydaz. Towards LLM-Powered Ambient Sensor Based Multi-Person Human Activity Recognition. *The 30th International Conference on Parallel and Distributed Systems*, Oct 2024, Belgrade, Serbia. <https://hal.science/hal-04619086v2>
- [24] Qu, H., Cai, Y., and Liu, J. LLMs are good action recognizers. *arXiv*. 2024. <https://arxiv.org/abs/2404.00532>
- [25] Ray, L. S. S., Zhou, B., Suh, S., and Lukowicz, P. Initial findings on sensor-based open vocabulary activity recognition via text embedding inversion. *arXiv*. 2025. <https://arxiv.org/abs/2501.07408>
- [26] Backman E, Granlund M, Karlsson AK. Parental Perspectives on Family Mealtimes Related to Gastrostomy Tube Feeding in Children. *Qual Health Res*. 2021;31(9):1596–1608. doi:10.1177/1049732321997133
- [27] Zhao, L., Garcia, C., Komizunai, S., Colley, N., Sato, A., Kouchiyama, M., Nasu, T., and Inoue, S. Nurse Activity Recognition in Gastrostomy Tube Feeding Using Video-Based Pose with Large Language Model-Guided Features. *International Journal of Activity and Behavior Computing*, 2025(1), pp. 1–28. <https://doi.org/10.60401/ijabc.38>
- [28] 小水内俊介, 二宮伸治, 金井理, 等. 気管内吸引, 体位変換および経管栄養の訓練が可能な生体反応を呈する看護教育シミュレータ [C]//ロボティクス・メカトロニクス講演会講演概要集 2021. 一般社団法人 日本機械学会, 2021: 2A1-E06.
- [29] 長内 真理乃ほか. 気管内吸引技術における動作分析を用いた習熟度評価の検討: 前屈姿勢, 手背移動, 停留, NASA-TLX を指標として. *日本小児呼吸器学会雑誌 = Japanese journal of pediatric pulmonology*. 30(2):2019,p.172-180.
- [30] 加納 敬ほか. 気管内吸引操作の量的評価—手首装着型 3 軸加速度センサの応用—[J]. *医療機器学*, 2022, 92(3): 255–262.