

発達障害者の異常行動認識における時間パラメータの最適化

藤岡 泰平¹ クリスティーナ ガルシア¹ 金子 晴¹ 井上 創造¹

概要：本研究では、発達障害者の異常行動を模擬したポーズ推定データに対し、行動の文脈情報を大規模言語モデル（LLM）に組み込むことで、時間パラメータ（ウィンドウサイズ、オーバーラップ率、シーケンス長）を最適化する手法を提案する。発達障害者支援施設においては、職員の人員不足や微細な動作の見落としが課題であり、異常行動の検出は困難を伴う。既存手法では、これらの行動が不規則かつ予測困難であるため、誤検出や見逃しが頻発していた。本研究の主な貢献は、異常行動を含む独自のラベル付きデータセットの構築と、Zero-shot および Few-shot プロンプトによる LLM の活用方法を示したことである。収集した行動データの文脈情報を活かして LLM にプロンプトを与えることで、各行動の特性に応じた最適なウィンドウサイズ、オーバーラップ率、LSTM の系列長を自動的に提案させた。データセットは、5 人の健常者が実演した 8 種の行動（うち異常行動 4 種）を記録したものであり、対象に障害のある方は含まれていない。評価実験では、全行動および異常行動ごとの分類性能について、LLM を用いないベースライン手法と比較した。その結果、Few-shot プロンプトを用いることで、「物を投げる」で 7.69 %、「攻撃する」で 7.31 %、「頭を振る」で 4.68 %、「爪を噛む」で 1.24 % の F1 スコアの改善が見られた。Zero-shot プロンプトにおいても、すべての異常行動で F1 スコア 96 % 以上を達成した。

Optimizing Time Windows for Recognizing Abnormal Behavior in Individuals with Developmental Disorders

Taihei Fujioka¹ Christina Garcia¹ Haru Kaneko¹ Sozo Inoue¹

1. はじめに

障害者施設におけるスタッフの不足は、施設運営の質やサービス提供に大きな影響を与える重要な問題である [1, 2]。福祉医療サービス機構（WAM）の調査によると、2023 年時点で施設の 52.6% がスタッフ不足を報告しており、2020 年の調査と比較して増加している [3]。知的障害や重度の精神疾患を持つ個人は、適切な診断とケアが提供されない場合、ケアの質の低下や安全性の問題に直面することがある。スタッフの診断能力やケアの公平性などの問題が特に顕著であり、これらの課題を解決するためには、多職種の協力と新技術の導入が必要である [4]。

現在、多くの障害者施設は、直接観察や手動記録に依存しており、これがスタッフの負担を増加させ、異常行動の正確な特定を難しくしている。また、手動記録には限界が

あり、十分なデータを継続的に収集することが困難である。データ収集がシステムによって行われている施設もあるが、それに伴い、システムの中断や誤ラベル、誤ったタイムスタンプといった問題が発生し、データ収集の精度や量に影響を与えることがある。これらのデータ収集の課題を克服し、異常行動検出の精度を向上させるために、映像情報を用いた人間活動認識技術が重要なアプローチとして研究されている。この技術は、異常行動の早期検出や日常活動との識別に有望ですが、ユーザが示す複雑で非典型的な行動の認識には課題が残る。人間が認識可能な異常の定義があいまいであるため、ビデオによる検出精度が制限される。これらの課題に対処するには、行動の文脈と異常行動認識のために特別に設計されたデータセットを考慮した新しいアルゴリズムの開発が必要である。

本研究では、短時間で発生する異常行動と、長時間にわたる日常的な行動の違いに対応するために、Zero-Shot および Few-Shot の手法を比較し、それぞれの特徴に応じた

¹ 九州工業大学
Kyushu Institute of Technology, Kitakyushu 808-0196,
Fukuoka, Japan

時系列の扱い方（パラメータ）を柔軟に生成することで、認識精度の向上を目指している。システムは、LSTMなどのモデルを使用して時間パラメータを調整し、行動を分類する。また、LLMを使用して独自のデータセットにウィンドウサイズ（時系列データを分割する1区間の長さ）、オーバーラップ率（各ウィンドウ間の重なり具合）、シーケンス長（LSTMに与える入力系列の長さ）を生成し、動作の前後関係や流れを考慮した、文脈理解に基づく高精度な認識モデルの構築を目指す。

提案されたシステムは、異常行動の認識において大きな改善を示した。結果は、Few-Shotのプロンプトが最良の認識精度を示し、モデルは爪を噛むで1.24%、頭を叩くで4.68%、物を投げるで7.69%、攻撃で7.31%のF1スコアの改善を示した。Zero-Shotのプロンプトも同様に高い認識精度を示し、すべての異常行動で96%以上のF1スコアを達成した。全体として、このシステムが現実の介護環境で異常行動を効率的に識別するのに役立つ可能性があることが示唆された[31]。さらに本研究では、時間的特徴量の多様性と複雑性に対応するため、複数の時間スケールや視点からの動作パターンを同時に捉えることを目的として、Parallel LSTM（並列LSTM）アーキテクチャを導入した。これは、異なるウィンドウサイズやシーケンス長を持つ複数のLSTMネットワークを並列に配置し、それぞれが異なる時系列的特徴を抽出・学習する構造である。この構成により、短期的な動作変化と長期的な行動の文脈を同時に考慮することが可能となり、特に異常行動と日常行動が混在する複雑な場面において有効であることが示された。実験結果からも、Parallel LSTMを用いたモデルは従来の単一LSTM構造と比較して、全体的な認識性能の向上がF1スコアで2.2%確認され、本提案手法の有効性をさらに裏付けるものとなった。

2. 関連研究

本節では、異常行動認識に関する既存研究を概観し、本研究で提案するフレームワークの動機となる課題を明らかにする。具体的には、異常行動の検出におけるコンピュータビジョンの一般的な手法、映像を用いた姿勢推定に内在する課題、ならびに認識精度向上に寄与する時間パラメータ最適化の重要性について論じる。さらに、人体行動認識（Human Activity Recognition: HAR）における大規模言語モデル（Large Language Models: LLMs）の活用可能性とその限界について検討する。最後に、LLMによって示唆された時間的パラメータを統合することで、上述の課題に対処する本研究の提案手法を提示する。

2.1 異常行動認識

近年、コンピュータビジョンおよび深層学習技術は、人間行動認識（Human Activity Recognition）分野において

広く応用されている[13]。特に異常行動の認識においては、大規模群集や施設環境における行動パターンの識別を目的とした研究が進められている。たとえば[14]では、新たに提案されたHAJJv2データセットに対し、手動でアノテーションされた異常行動を対象に、CNNとランダムフォレスト（RF）を組み合わせた手法が用いられている。具体的には、少人数の群集にはResNet-50を、大規模群集にはYOLOv2を適用し、データの規模や特性に応じた異常行動の認識を行っている。ここで、小規模群集は数十人規模、大規模群集は数百人から数千人に及ぶ群衆を指し、遮蔽や動きのブレ、行動の多様性といった課題が顕著である。同研究ではAUC 76.08%を達成し、新たな基盤を提示したが、群集の規模によってモデルを切り替える必要があるため、個別行動や複雑な動作の認識には限界がある。これに対して本研究では、骨格推定に基づく手法を採用し、個人単位での高精度な認識と環境の多様性への柔軟な対応を実現するフレームワークを提案する。

2.2 映像を用いた姿勢推定と行動認識

異常行動認識においては、RGB映像および骨格情報を用いた手法が広く用いられている。特に骨格ベースの姿勢推定は、映像データに比べて処理が簡易であり、効果的であるとされている。しかし、LSTM（Long Short-Term Memory）モデルは空間情報の抽出に限界があり、次元削減や時系列データの解析において課題が指摘されている[16]。加えて、作業環境や対象者の身体的特徴によって生じるデータの不均衡は、モデルの精度を著しく低下させる要因となっている[16]。

一方で、骨格情報に基づく手法は、照明条件や背景の違いに対して頑健であり、RGBベースの手法と比較して環境変化に対する適応性が高いことが報告されている[26]。さらに、骨格データは外見情報に依存しないため、カメラアングルや衣服の違いに左右されにくく、CNN（畳み込みニューラルネットワーク）やオブティカルフローに基づく手法よりも安定した動作認識が可能である[27]。これらの特性から、本研究では骨格ベースの特徴量が異常行動認識に適していると判断した。

このような背景のもと、[17]では、骨格データにおける空間的および時間的な関係性を効果的に捉えるため、Spatio-Temporal Graph Convolutional Networks（ST-GCN）と正規化アテンション機構（Normalized Attention Mechanism: NAM）を組み合わせた手法が提案されている。この手法は、異常行動の特徴抽出精度を向上させることに成功しているが、短時間で発生する行動や文脈的に曖昧な行動の認識には依然として限界がある。そこで本研究では、こうした課題に対処すべく、当該行動に特化したパラメータ設計により認識性能の向上を図ることを目的としている。

2.3 異常行動における時間パラメータ最適化

時系列データの解析において、時間パラメータの設定は極めて重要な要素であり、特に LSTM モデルを用いた異常行動認識においては、その最適化が課題となっている。たとえば、EEG データを用いた研究 [18] では、LSTM における複数の時間パラメータを評価し、分類精度を最大化する最適な時間パラメータを選択する手法が提案されている。この研究では、時間パラメータの選定がデータの特性に大きく依存することが示されており、異常行動認識においても、対象行動やデータ分布に応じて動的に最適な時間パラメータを選定する必要性が示唆されている。

また、周期性や長期依存性を考慮した可変長の時間窓を用いた手法も提案されており、こうした手法がデータ解析の有効性を高めることが報告されている [19]。これらの知見は時間パラメータ最適化の重要性を支持するものであるが、現実環境への適用については未だ十分に検証されていない。

2.4 人体行動認識における大規模言語モデルの応用

近年、人体行動認識 (Human Activity Recognition: HAR) における大規模言語モデル (Large Language Models: LLMs) の応用が急速に進展している。従来の手法と異なり、LLM はセンサデータや時系列情報を直接処理可能であり、Zero-shot 学習やプロンプト設計を通じて効率的かつ高精度な行動分類を実現する。

たとえば、Ji ら [20] は、加速度・ジャイロデータを GPT-4 で処理する HARGPT を提案し、Capture24 および HHAR といったデータセットにおいて 80% を超える分類精度を達成した。これに対し、Shao ら [28] の TCN-attention-HAR は、UCI HAR データセットを用いて手工芸特微量と時系列畳み込みネットワークにより 95.69% の精度を報告している。

さらに、Shoumi と Inoue [11] は、推論過程を逐次的に説明する Chain-of-Thought (CoT) プロンプティングの導入により、モデルの透明性と解釈性の向上を図っている。これにより、ユーザがモデルの判断根拠を理解しやすくなり、信頼性と実用性が高まる。

LLM はまた、センサ情報を自然言語表現へ変換することでマルチモーダルデータの統合にも効果的であり、手作業による特微量設計を不要にすることで、異常行動と日常行動の柔軟な分類を可能にする [16]。特に Shoumi と Wei [21] は、複雑な動作パターンや文脈情報の認識における LLM の有効性を示し、高精度な分類性能を達成している。一方で、LLM の計算コストやデータフォーマットの整合性への依存といった課題も依然として存在しており、今後の研究ではプロンプト設計の最適化やモデル効率の改善が求められる。

本研究ではこれらの知見を踏まえ、骨格特微量に対して

動的な時間パラメータ最適化メカニズムを統合することで、短時間かつ複雑な異常行動の高精度な認識を目指す。正常参加者による異常動作の模擬データを収集し、YOLOv7 を用いて映像から姿勢情報を抽出した。YOLOv7 は、リアルタイム処理において精度と速度のバランスに優れており、YOLOR、YOLOX、YOLOv5、DETR などと比較して広範な FPS 範囲において優れた性能を示すため本研究で採用している [22]。収集したデータを用いて、Zero-shot および Few-shot のプロンプトから生成された時間的パラメータを比較・評価し、最適な文脈と事例による認識精度の向上を図った。

3. 健常者による模擬行動から構築したデータセット

本セクションでは、異常および正常行動を記録した映像に対して適用した姿勢推定手法を含めた、データ収集のプロセスについて詳述する。

3.1 実験室内におけるデータ収集

本研究で使用したデータセットは、実験室内で制御された環境下において、延べ 5 名の健常参加者を対象に、4 日間にわたって収集されたものである [30]。撮影には 2 台の異なるカメラを使用し、記録された映像に対して姿勢推定を適用することで、骨格データを抽出した。

対象とした行動は、異常行動 4 種 (例: 「物を投げる」「頭を振る」など) および正常行動 4 種 (例: 「歩行」「座る」など) から構成されており、表 1 に一覧を示す。これらの行動の選定は、発達障害者支援施設において実施された観察およびインタビューに基づいて行われた。施設職員によれば、これらの行動は頻繁に観察され、利用者の安全確保や支援の質を高めるうえで注意を要するとのことであった。

表 1: 対象行動と行動の種類

行動	行動分類
静かに座る	正常
スマートフォン使用	正常
歩く	正常
間食	正常
頭を振る	異常
物を投げる	異常
攻撃する	異常
爪を噛む	異常

本研究では、発達障害者支援施設における異常行動認識を目的とし、計画的なデータ収集実験を実施した。対象行動は、自傷・攻撃的な異常行動 (頭を振る、物を投げる、他者を攻撃する、爪を噛む) と、日常的な正常行動 (座る、歩く、スマートフォン使用、間食) で構成され、実環境での応用を想定した内容となっている。

参加者は20～30代の健常な大学院生5名であり、各自60分×4回のセッション（計240分）において、指定された8行動を時間制限付きで再現した。撮影はGoPro9およびiPhone15Proを用いて複数角度から行い、得られた映像から骨格キーポイント（頭部、肩、手首、膝など）を抽出し、Vrewにより全フレームを手動ラベリングした。

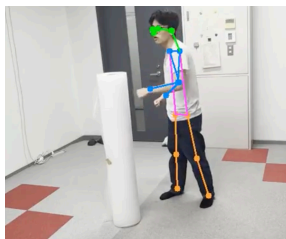
実験は既存カメラ配置で行い、追加設置は不要と判断された。すべてのデータは匿名化され、健常者による模擬動作を対象としている。収集データは行動認識モデルの学習および精度評価に用いられた。

3.2 行動分布と姿勢推定

本研究では、参加者の動作をGoPro 9およびiPhone 15 Proにより30fpsで記録し、姿勢推定手法を用いて骨格データを抽出した。映像処理にはYOLOv7を用い、鼻・目・耳・肩・肘・手首・腰・膝・足首の17箇所に対応するキーポイントを取得した。

図1は再現された異常行動の例を示すもので、(a) 攻撃する、(b) 爪を噛む、(c) 頭を振る、(d) 物を投げるに対応する。攻撃や物を投げるなどの高エネルギー動作は大きな関節の明瞭な動きが特徴であり、モデルによって比較的認識しやすい。一方、爪を噛む・頭を振るなどの微細かつ反復的な動作は、検出が難しくなる傾向が見られた。

これらの結果は、動作の強度やスケールに応じた姿勢推定モデルの調整の必要性を示している。今後は、行動ごとに重要な関節へ重みを付与する手法や、時間的な軌跡情報を活用することで、特に反復的な異常行動の認識精度向上が期待される。



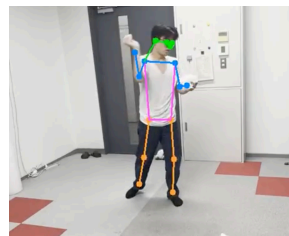
(a) 攻撃する



(b) 爪を噛む



(c) 頭を振る



(d) 物を投げる

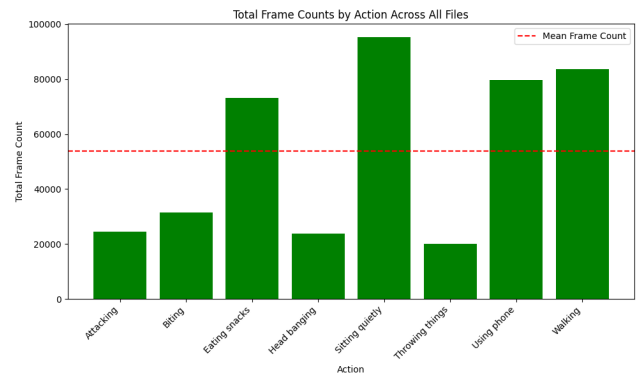
図1: 健常な被験者が実施した異常動作

図2は、各行動クラスにおける総フレーム数を可視化し

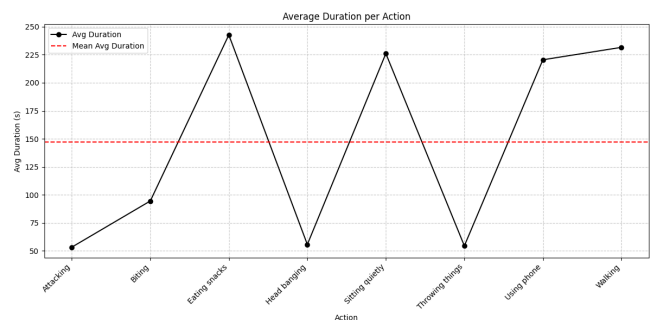
たものである。この図は、異常行動および通常行動を含む8つの行動クラスの分布を示しており、データセットのバランスを評価し、認識モデルの性能に影響を与える可能性のあるバイアスを特定するうえで重要である。例えば、動作ごとのフレーム数に偏りがあると、学習時に特定のクラスが過剰に適合（過学習）したり、逆に十分に学習されなかったりする可能性があり、モデルの汎化性能に悪影響を与える。

図2(a)は、各行動クラスの総フレーム数を示しており、「静かに座る」や「歩行」といった通常行動が他の行動より長時間持続し、データ量が多いことが確認できる。赤い破線は平均フレーム数を示し、各クラスの相対的な分布を視覚化している。

図2(b)は、平均継続時間を示しており、「おやつを食べる」「静かに座る」は長く、「攻撃」「物を投げる」は短い傾向が見られる。これにより、動作ごとの時間的特徴の違いが明らかとなり、学習時の考慮が必要であることが示唆される。



(a) 行動カテゴリ別のフレーム数



(b) 行動カテゴリ別の平均継続時間

図2: 全ファイルにおける行動カテゴリ別のデータ分布と継続時間

これらの結果は、総フレーム数や継続時間の違いがモデル学習に影響を与えることを示している。通常行動が多すぎると異常行動の分類精度が低下する可能性があり、また継続時間のばらつきは時間的感度の重要性を示唆する。ウィンドウ調整や重み付きサンプリングの導入により、よ

り公平で頑健な学習が期待される。

4. LLM を用いた異常行動認識手法の提案

図 3 は、本研究で提案する姿勢推定と LLM による時間パラメータ最適化を組み合わせた異常行動認識のワークフローを示している。映像は正面カメラで撮影され、YOLOv7 Pose Estimation により骨格キーポイントが抽出される。前処理では、人物の追跡や NMS による信頼性の高いキーポイント選別を行い、時間的整合性のある構造化データを生成する。その後、LLM によりウィンドウサイズやオーバーラップ率などの時間パラメータを最適化する。

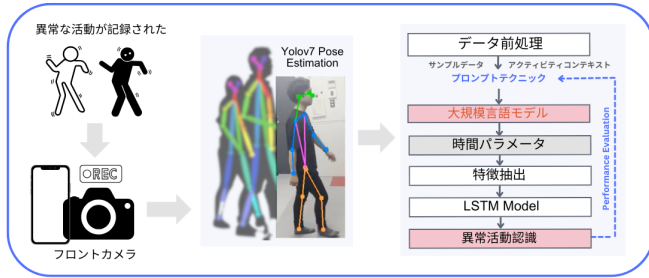


図 3: 異常行動認識フレームワークの概要

本データセットは時系列動作データを含むため、時間的依存関係を効果的に捉えるために LSTM モデルを採用する。LSTM はゲーティング機構を通じて長期的な情報保持が可能であり、勾配消失問題を軽減する特徴を持つ [23]。また、Transformer のようなアテンションベースのモデルと比較して、LSTM は計算資源の消費が少なく、小規模なデータセットにおいて過学習のリスクも低い [24]。さらに、LLM は時間パラメータの最適化を支援し、従来のような広範な手動チューニングの必要性を軽減する [25]。

最適化されたデータは特徴量抽出を経て、活動パターンの学習および分類を行う LSTM モデルに入力される。本システムは異常行動の検出を行い、その精度および信頼性に基づいて評価される。最後に、異なるプロンプト手法によって得られた時間パラメータによる性能比較を行う。

4.1 時間パラメータ生成におけるプロンプト手法

図 4 は、異常行動の認識に必要な時間的パラメータ（ウィンドウサイズ、オーバーラップ率、シーケンス長）を生成する際に用いた、Zero-Shot および Few-Shot プロンプト手法の違いを示している。Zero-Shot では、あらかじめ基準となるパラメータを提示し、動作の特徴に基づいて直接的に最適化を指示する。

一方、Few-Shot では、ラベル付きの時系列特徴量データを提示し、データの内容や文脈を踏まえて柔軟に最適化目標を構築する。これにより、対象動作やデータ構造に応じたパラメータ生成が可能となる。本手法は、現実の行動認識タスクにおいて、柔軟かつ汎用的な時間的処理の設計を

支援する。

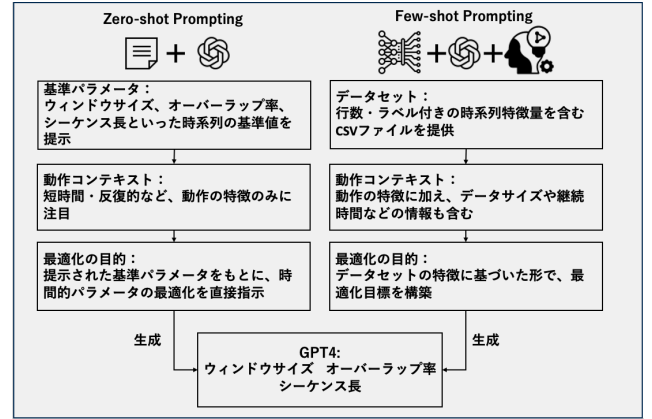


図 4: 時間パラメータ最適化のためのプロンプト設計

一般に、LLM に時間パラメータを生成させるプロンプト処理は、式 1 のように数学的に表現できる。ここで、 T は式 2 で詳述される時間パラメータを表し、 P はプロンプト手法を示す。

$$T = \arg \max_T p_G(T | P) \quad (1)$$

$$T = \text{ウィンドウサイズ, オーバーラップ率, シーケンス長} \quad (2)$$

入力プロンプト P の種類は、使用するプロンプト手法によって異なる。Zero-Shot プロンプトでは、 P_{ZeroShot} への入力には、活動の説明、ベースラインパラメータ、活動の文脈、および最適化目標が含まれ、これは式 3 で定義される。一方、Few-Shot プロンプトでは、式 4 に示すように、これらに加えてデータセット固有の具体的な例が含まれる。

$$P_{\text{ZS}} = \{T, P, C, O\} \quad (3)$$

$$P_{\text{FS}} = \{D, T, C, O\} \quad (4)$$

これらの定式化により、LLM は活動やデータセットの特性に適応した時間パラメータを自動生成でき、本研究ではこの特性を活かして柔軟なモデル設計を行っている。従来手法では難しい最適化に対し、LLM は自然言語プロンプトを通じて動作の文脈を反映した柔軟な最適化が可能である。Zero-Shot プロンプトでは、事前のデータなしに、ベースライン設定や動作特性、最適化目標を LLM に与えることで、時間パラメータ（ウィンドウサイズやシーケンス長など）を自動的に導出する。使用したテンプレートは以下に示す。

- 対象行動 (T) : 認識対象の異常行動と、使用する認識モデルを明記する。
- ”ベースラインパラメータ (P) ”: ウィンドウサイズ、オーバーラップ率、シーケンス長など、時間パラメータ

タの初期設定を記述する。これらは最適化の出発点となる。

- 行動の文脈（*C*）：対象行動の特性（動作の種類、反復性、リズムなど）を記述し、行動に応じた最適化を行う。たとえば、「攻撃する」は短く速い動き、「爪を噛む」や「頭を振る」は反復的かつリズムカルな動作、「物を投げる」は明確な腕の動きを伴う短時間の動作に注目する。
- 最適化目標（*O*）：行動認識性能の向上と計算効率の両立など、最適化の目的を定義する。

一方、Few-Shot プロンプトでは、ラベル付きの時系列データを含む CSV ファイルなど、データセット固有の特徴を取り入れている。これにより、LLM はデータセットに特化した情報を考慮しながら最適化目標を設定することが可能となる。どちらの手法も、LLM が活動ごとの時間パラメータを効率的に導出する能力を活用しており、計算効率を考慮しつつ異常行動を高精度に認識できるようにしている。Few-Shot 用のテンプレートは以下のとおりである：

- ”対象行動（*A*）”：同様に、認識対象の異常行動を指定する。
- ”データセット（*D*）”：分析に使用するデータセットの詳細を記述する。ファイル形式、データポイント数、行動ラベルを含む例示データの情報などを含む。
- ”行動の文脈（*C*）”：動作の特性に加え、すべての動画における当該行動の合計時間など、データセット固有の情報を含む。
- ”最適化目標（*O*）”：認識精度と効率性のバランスを考慮し、データセットの特性に基づいた最適化目標を定義する。

本研究では、異常行動認識における時間パラメータ生成に GPT-4-turbo を使用した。temperature は出力の多様性と一貫性を両立させるため 0.7 に設定し、top-p は 0.9 とすることで自然な出力を維持した。文脈を十分反映させるため、最大トークン数は 25,000 とし、frequency penalty および presence penalty は出力の安定性確保のため 0 に設定した。これらの値は事前実験により最適と判断された。

4.2 2D 姿勢推定からの特徴量抽出

本研究では、右手・左手・肩・目・足など計 8 点のキーポイントを用い、対象行動に関連する部位の組み合わせから 14 種の特徴量を算出した。これらは動作や空間的特徴を捉え、異常行動認識の精度向上に寄与する。

2D 姿勢推定を採用したのは、RGB や深度ベース手法に比べて計算効率が高く、ノイズに強いためである。スケルトンデータは動作の本質を簡潔に表現でき、環境の影響を受けずに空間的・動的な行動分析が可能となる。

速度と加速度は、右手・左手および右足・左足といった主要な身体部位に対して取得され、動作中の速度（ v , Equation 5）および加速度（ a , Equation 6）を測定する。これらの特徴量は、異常行動の認識に重要な動的かつ急速な動きのパターンを強調する。角度（ θ , Equation 7）は、右肩と右手首、左肩と左手首といった主要な関節間で算出され、四肢の位置関係や姿勢を捉えるために用いられる。眼球の動きは、右目と左目の垂直方向の変位（ Δy_{eye} , Equation 8）および水平方向の変位（ Δx_{eye} , Equation 9）を通じて分析され、視線方向や動作パターンに関する手がかりを提供する。最後に、距離特徴量（ d , Equation 10）は、右手と右目、または左手と左目といった特定のキーポイント間の空間的關係を測定し、身体部位間の相対的な位置関係を追跡する。これらの特徴量は、動作や空間的な変化を総合的に捉え、異常行動の認識精度を高める。加速度は「攻撃する」のような急激な動作に有効であり、速度やリズム指標は「頭を振る」「爪を噛む」といった反復的な行動の識別に寄与する。

$$v = \frac{\|\mathbf{p}(t) - \mathbf{p}(t-1)\|}{\Delta t} \quad (5)$$

$$a = \frac{v(t) - v(t-1)}{\Delta t} \quad (6)$$

$$\cos \theta = \frac{(\mathbf{p}_B - \mathbf{p}_A) \cdot (\mathbf{p}_C - \mathbf{p}_B)}{\|\mathbf{p}_B - \mathbf{p}_A\| \cdot \|\mathbf{p}_B - \mathbf{p}_A\|} \quad (7)$$

$$\Delta y_{eye} = y_{eye}(t) - y_{eye}(t-1) \quad (8)$$

$$\Delta x_{eye} = x_{eye}(t) - x_{eye}(t-1) \quad (9)$$

$$d = \|\mathbf{p}_A - \mathbf{p}_B\| = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \quad (10)$$

4.3 異常行動認識

本研究では、異常行動認識のために時系列依存性の学習に適した LSTM モデルを採用し、30fps で取得した詳細な時系列データを、固定長の時間窓で分割した。ウィンドウサイズやオーバーラップ率は精度と計算効率の両立を目指して最適化された。

評価には 5 名のデータを用いた Leave-One-Person-Out クロスバリデーション（LOPO-CV）を採用し、未知の人物への汎化性能を検証した。LSTM ユニット数は 64、エポック数は 20、バッチサイズは 32 とし、各ウィンドウに対して精度、適合率、再現率、F1 スコアを記録し、混同行列を集約して全体性能を評価した。

5. 発達障害者における行動分類の結果

本研究では、姿勢推定データを活用して異常および正常な人間行動を認識する映像ベースのフレームワークを提

案・評価した。評価は、5名の参加者に対し、1名をテスト対象とする Leave-One-Subject-Out 法を用いて 5 分割交差検証を行い、正常行動と異常行動の認識性能を比較・分析した。

本節では、本フレームワークを用いたデータセットの実装結果を示す。LSTM により行動の時間的依存性を捉え、ウィンドウサイズ・オーバーラップ率・シーケンス長を調整することで、異常行動と正常行動の両方で高精度な認識が可能となった。

これらのパラメータ調整により、短時間の異常行動を正確に認識しつつ、正常行動の高精度な認識も維持することが可能となり、本フレームワークの適応性とケア・モニタリング応用における有用性が示された。特に短時間の異常行動の精密な検出を実現しつつ、正常行動に対する高い認識精度も確保されており、本手法が実環境での応用において高い可能性を有することが実証された。

5.1 全行動に対する認識結果

本節では、4 種類の正常行動および 4 種類の異常行動すべてに対する認識性能の結果について、時間パラメータの違いに着目して詳述する。

これらの行動を認識するために、LSTM モデルを用いて時系列データを処理した。ウィンドウサイズ、オーバーラップ率、シーケンス長といった設定を変更しながら、それぞれのパフォーマンスを比較した。各設定の詳細およびその結果は以下に示す。

表 2: 異なる時間パラメータ設定における認識性能

アプローチ	時間パラメータ			認識性能 (%)	
	WS	OR	SL	Accuracy	F1 スコア
A	30	50%	30	71	69
B	30	50%	60	72	70
C	30	50%	90	75	76
D	30	50%	120	74	75
E	30	50%	150	75	75

モデル性能の評価においては、5つの異なる設定（アプローチ A、B、C、D、E）を採用した。これらの設定は、時間的文脈が認識性能に与える影響を分析することを目的として設計されており、ウィンドウサイズを 30 フレーム、オーバーラップ率を 50%に固定したうえで、シーケンス長のみを 30 フレームから 150 フレームまで 30 フレーム刻みで変更した。

その結果、シーケンス長を 90 フレームまで増加させることで性能が向上し、この設定において最高の認識精度 75%、F1 スコア 76%を達成した。しかし、90 フレームを超えてシーケンス長をさらに増加させても性能の向上は見られなかった（表 2 参照）。特に、90 フレームを超えると認識精度はほぼ横ばいとなり、時間的文脈を過度に拡張し

ても認識性能のさらなる改善には必ずしもつながらないことが示された。

5.1.1 Parallel LSTM の認識結果

本研究では、短期および長期の文脈情報を同時に捉えることを目的として、Parallel LSTM モデルによるマルチスケール時系列処理を導入した（表 3）

表 3: Parallel LSTM の認識性能

アプローチ	時間パラメータ			認識性能 (%)	
	WS	OR	SL	Acc.	F1
A	30	50%	90&15	74	74
B	30	50%	90&30	73	73
C	30	50%	120&15	77	77
D	30	50%	120&30	75	75
E	30	50%	150&15	72	72
F	30	50%	150&30	76	77

長時間の動作（90～150 フレーム）をそれぞれ個別に処理する 2 系列の LSTM を並列に構成し、両者の出力を結合することで、瞬時的かつ持続的な行動パターンを同時に捉えるアーキテクチャを設計した。なお、ウィンドウサイズは一律 30 フレーム、オーバーラップ率は 50%に固定し、短期シーケンス（15 フレーム）長に対する長期シーケンス長を 90、120、150 フレームに変更して性能を比較した。

その結果、長期シーケンス長が 120 フレーム、短期シーケンス長が 15 フレーム（120&15 frames）の構成において、最も高い認識性能（Accuracy: 77%、F1 スコア: 0.77）を達成した。この結果は、単一スケールの LSTM を用いた従来手法（5.1 節）における最良性能（Accuracy: 75%、F1 スコア: 0.76）を上回っており、複数の時間スケールを統合的に処理するアプローチの有効性を示すものである。一方で、長期シーケンス長を 150 フレームに拡張した構成では性能向上は確認されず、過度に長い時系列入力モデルの学習効果を低下させる可能性も示唆された。

5.2 各行動ごとの認識結果

本節では、各異常行動に対して全正常行動を比較対象とし、最適化された時間パラメータが認識性能に与える影響を評価する。本フレームワークは、異常行動が短時間のみ通常行動に混在する現実的なシナリオを想定して設計されている。

検証では、「攻撃する」「爪を噛む」などの異常行動を個別に評価し、他のすべての行動を正常として扱って識別性能を比較した。Zero-shot および Few-shot の 2 手法に基づく評価結果は、表 4 および表 5 に示す。

Zero-shot と Few-shot は、それぞれ異なる利点と制約を持つ。Zero-shot はラベル不要で迅速に推定できるが、複雑な行動の認識には限界がある。一方、Few-shot は少量の学習データで精度を向上できるが、追加データの準備が必

表 4: Zero-shot 学習における行動認識性能の分析

対象異常行動	時間パラメータ			認識性能 (%)	
	WS	OR	SL	Acc.	F1
正常 vs. 攻撃する	20	50%	40	98.50	96.27
正常 vs. 爪を噛む	15	66%	40	99.49	98.81
正常 vs. 頭を振る	30	66%	60	99.23	97.63
正常 vs. 物を投げる	20	66%	60	99.61	98.62

要となる。Zero-shot は初期分析向き、Few-shot は高精度な最適化に有効だが、どちらも万能ではない。

本研究では、Zero-shot では LLM から得たパラメータのみを用いて行動を反映させ、Few-shot では少量の学習データを用いて各行動に応じた調整を行い、精度向上を図った。

5.2.1 Zero-shot 結果の性能分析

Zero-shot シナリオにおける認識性能を分析するため、各異常行動と正常行動を比較して評価を行った。Zero-shot 学習では、テキストベースのプロンプトのみを用いて LLM から得られた事前学習知識に基づき、最適な時間パラメータを推定する。本評価では、異常行動をどの程度正確に識別できるかを明らかにするとともに、特定の異常行動における認識の課題を検討した。結果は表 4 にまとめられている。

Zero-shot の結果から、異常行動ごとの認識性能に差が見られた。「爪を噛む」は F1 スコア 98.81% と最も高く、リズムミカルな動作が識別に有効だった。一方、「攻撃する」は 96.27% と最も低く、正常行動との類似性が誤分類の要因と考えられる。

5.2.2 Few-shot 学習における行動認識性能の分析

表 5: Few-shot における異常行動認識性能

対象異常行動	時間パラメータ			認識性能 (%)	
	WS	OR	SL	Acc.	F1
正常 vs. 攻撃する	20	50%	80	99.09	97.48
正常 vs. 爪を噛む	20	50%	60	100.0	100.0
正常 vs. 頭を振る	20	75%	60	99.95	99.90
正常 vs. 物を投げる	25	80%	75	99.79	99.40

Few-shot 学習は、各行動に対してモデルのパラメータをさらに精緻化するために用いられた。少量のラベル付きデータを活用することで、ほとんどの行動においてほぼ完全に近い認識性能を達成しており、その結果は表 5 に示されている。

この手法では、実験の一貫性を保つために使用する特徴量サンプルの数を一定に設定した。また、Few-shot 学習の補強情報として、各行動の特徴量 CSV ファイルに加え、常に 2 枚の画像を併用した。具体的には、各行動シーケンスの先頭 2 枚の画像を選択し、処理の一貫性を確保した。

Few-shot 学習は、「攻撃する」や「物を投げる」といった難易度の高い行動において F1 スコアと正解率を大きく向

上させた。「爪を噛む」では両指標ともに 100.0% を達成し、提案パラメータの有効性が示された。一方で、「物を投げる」はばらつきが大きく、安定した分類には課題が残った。

発達障害者支援施設では、倫理的制約や発生頻度の低さから大量のラベル付きデータが得にくいため、少数データでも高精度を維持できる Few-shot 学習は有効であり、本手法が現場での応用に適していることが示唆された。

6. LSTM モデルにおける時間パラメータ設定の有効性の検討

本節では、提案フレームワークによって得られた結果を分析し、異常および正常行動の識別能力に着目して評価を行う。全体的な認識性能および行動別の認識性能を比較・分析することで、認識精度に影響を与える要因、特に時間パラメータ設定の有効性を明らかにすることを目的とする。

6.1 アプローチごとの分類性能と時間的文脈の効果

正常行動は高い精度で分類された一方、異常行動では一部誤分類が見られた。これは行動の類似性やデータの不均衡が影響した可能性がある。アプローチ A では短い依存関係の限界から誤分類が多く、アプローチ B・C では時間的文脈を拡張することで認識精度が向上した。本研究では単一の時間ウィンドウを使用した。今後は短期・長期ウィンドウを併用するマルチスケール手法が、異常行動の検出精度向上に有効であると考えられる。

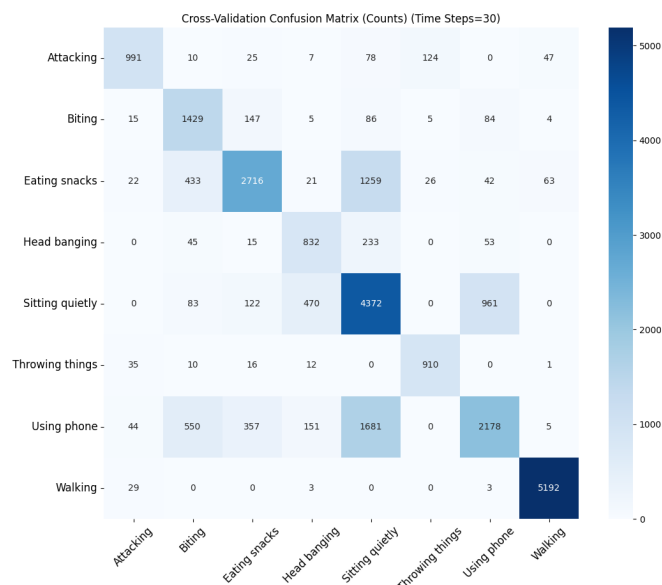


図 5: Confusion Matrix of Baseline, Approach A

図 5 は、30 タイムステップでの混同行列を示しており、「歩行 (Walking)」や「物を投げる (Throwing things)」は特徴的な動作により高精度に分類された。一方で、「スマー

トフォン使用 (Using phone)」や「間食 (Eating snacks)」は「静かに座る (Sitting quietly)」や「爪を噛む (Biting)」との誤分類が多く、類似姿勢の識別にはさらなる特徴量の工夫が必要である。

図 6 は、60 タイムステップでの混同行列を示しており、「歩行」や「物を投げる」は高精度に分類された。一方で、「スマートフォン使用」や「間食」は他行動との誤分類が依然多く、時間ウィンドウを長くしても細かな動作の識別には限界があることが示唆された。

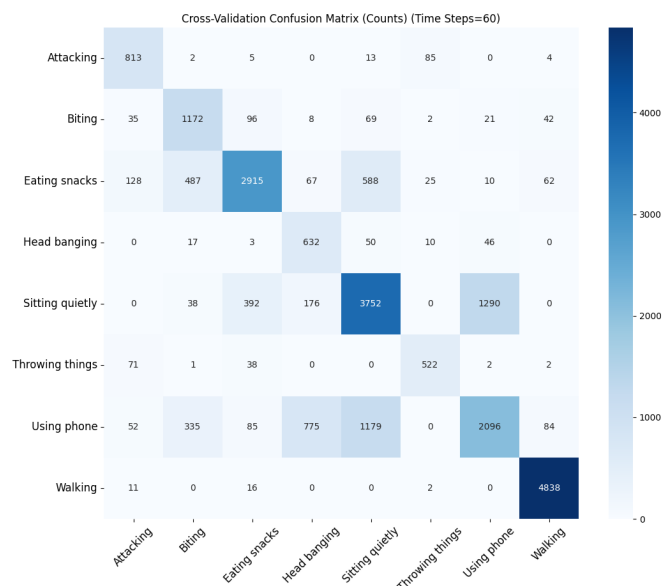


図 6: Confusion Matrix of Baseline, Approach B

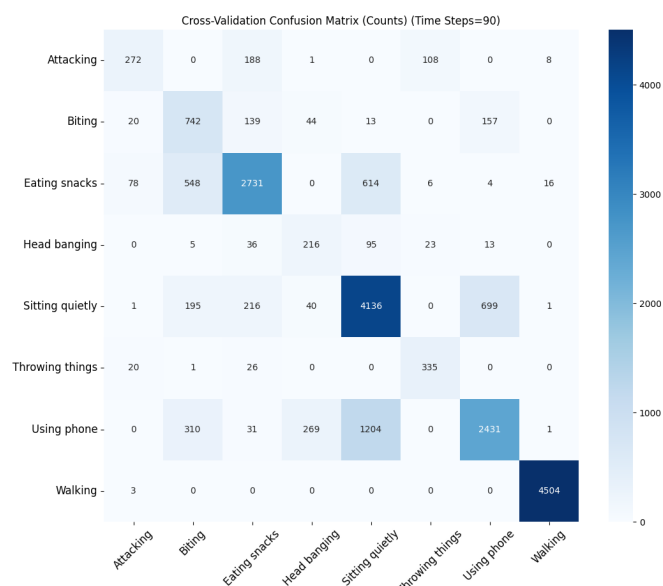


図 7: Confusion Matrix of Baseline, Approach C

図 7 は、90 タイムステップでの混同行列を示し、全体で最も高い認識精度を達成した結果である。「歩行」や「物を

投げる」は高い正解率を示した一方で、「スマートフォン使用」や「間食」は類似姿勢の行動と混同されやすく、識別に課題が残ることが示された。

これらの結果から、時間的ウィンドウを長くしても微細な動作の認識には限界があることが明らかとなり、特に手の動きが繊細であったり、姿勢が類似していたりする行動を正確に識別するには、今なお課題が残されている。

6.2 LLM を用いた時間パラメータの最適化に関する考察

本研究の結果は、提案フレームワークの柔軟性と実用性を示しており、動的な環境における異常行動の高精度な認識に有効である。

表 6: 正常行動 vs. 異常行動における認識性能

異常行動	プロンプト	F1 スコア (%)	改善幅 (%)
攻撃する	Baseline	90.17	
	Zero-shot	96.27	6.10
	Few-shot	97.48	7.31
爪を噛む	Baseline	98.76	
	Zero-shot	98.81	0.05
	Few-shot	100.00	1.24
頭を振る	Baseline	95.22	
	Zero-shot	97.63	2.41
	Few-shot	99.90	4.68
物を投げる	Baseline	91.71	
	Zero-shot	98.62	6.91
	Few-shot	99.40	7.69

表 6 に示すように、「爪を噛む」のような手の位置が一定である行動は Zero-shot でも高い識別性を示し、一方で「攻撃する (Attacking)」や「物を投げる」などの複雑な動作には Few-shot 学習が有効だった。今後は、Zero-shot と Few-shot を組み合わせたハイブリッド手法により、限られたラベル付きデータ環境でも高い認識精度と汎化性能を両立できる可能性がある。

本研究では、Few-shot プロンプトが最も高い認識性能を示し、ベースラインに対して F1 スコアが最大で 7.69% 改善された。また、Zero-shot でも全行動で F1 スコア 96% 超を達成した。これにより、本システムは介護現場における異常行動の効率的な検出支援に有効である可能性が示された。

7. 結論

本研究では、発達障害者を対象とした模擬的な異常行動に対する姿勢推定データを用いて、行動の文脈情報を大規模言語モデル (LLM) に取り入れることで、時間パラメータの最適化を図る手法を提案した。

本研究の結果、提案フレームワークは正常・異常行動のいずれに対しても高い認識精度を示した。特に、異常行動

の文脈に基づく時間パラメータの導入により、従来の単一フレーム手法を上回る性能が得られた。また、ウィンドウサイズやオーバーラップ率の調整によって長期的な時間文脈が捉えられ、行動の連続性の把握に有効であることが示され、時系列データの重要性が再確認された。

提案フレームワークは正常・異常行動のいずれにも高い認識精度を示し、特に「爪を噛む」は周期的な動作特性により最も高精度に分類された。一方、「攻撃する」は短時間かつ多様な動作により誤認識が生じやすく、行動ごとの時間パラメータ最適化の重要性が示された。

本研究の主な貢献は、異常行動にラベル付けを行った独自データセットの構築と、そのデータセットに対して Zero-shot および Few-shot プロンプトによる LLM 活用を比較・検討した点にある。提案手法では、収集した異常行動データの文脈をもとに、LLM に対してウィンドウサイズ、オーバーラップ率、LSTM モデルのシーケンス長といった時間パラメータの提案を促し、各行動の特性に適した設定を得ることが可能となった。

さらに、異常行動を迅速かつ正確に検出できる能力は、施設内の安全性向上のみならず、介護従事者の負担軽減にも寄与する。特に、正常行動の中に一時的に出現する短時間の異常行動を確実に識別できる点は、リアルタイムでの対応が求められる現場において重要な特徴である。

今後の課題としては、異常行動の多様性や個人差を考慮したデータセットの拡充が挙げられる。また、行動の文脈をより深く捉える時系列モデルの最適化も、さらなる性能向上につながると考えられる。加えて、他の行動認識アルゴリズムやフレームワークとの比較評価を通じて、本手法の実用性および汎用性をより明確にしていく必要がある。

本研究の成果は、行動モニタリング技術の有用性を示すものであり、発達障害者支援施設における安全性および業務効率の向上に向けた実践的な基盤を提供するものである。今後は、これらの知見をもとに、より高度かつ柔軟な異常行動認識システムの構築を目指して研究を継続していく。

参考文献

- [1] Spreat, S., and Conroy, J. A Supply and Demand Perspective on the Workforce Crisis in Intellectual Disability. *Journal of Intellectual and Developmental Disability*, pp. 1–5, 2021. doi: 10.3109/13668250.2021.1975365
- [2] Peters, V., Frielink, N., van Leest, C., Heerkens, L., and Embregts, P. J. C. M. Impact pathways: Putting workers front and center in addressing workforce shortages in intellectual disability care. *International Journal of Operations and Production Management*, 44(13), pp. 251–262, 2024. doi: 10.1108/IJOPM-03-2023-0175
- [3] Japan Welfare and Medical Service Agency. Survey on securing human resources for disability welfare services, etc. in 2023. Research Group, Management Support Center, 2024.
- [4] Ashok, N., Hughes, D., and Yardley, S. Challenges and opportunities for improvement when peo-

- ple with an intellectual disability or serious mental illness also need palliative care: A qualitative meta-ethnography. *Palliative Medicine*, pp. 1–16, 2023. doi: 10.1177/02692163231175928
- [5] Ke, S.-R., Le, H. U. T., Yoo, J.-H., Lee, Y.-J., Hwang, J.-N., and Choi, K.-H. A Review on Video-Based Human Activity Recognition. *Computers*, 2(2), pp. 89–93, 2013. doi: 10.3390/computers2020089
- [6] Inoue, S., Lago, P., Hossain, T., Mairittha, T., and Mairittha, 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, pp. 1–24, 2019. doi: 10.1145/3351244
- [7] Garcia, C., and Inoue, S. Challenges and Opportunities of Activity Recognition in Clinical Pathways. In **Human Activity and Behavior Analysis**, CRC Press, pp. 1–18, 2024. doi: 10.1201/9781003371540-8
- [8] Morais, R., Le, V., Tran, T., Saha, B., Mansour, M., and Venkatesh, S. Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11988–11996, 2019. doi: 10.1109/CVPR.2019.01227
- [9] Fikry, M., Garcia, C. A., Vu, Q. N. P., Inoue, S., Oyama, S., Yamashita, K., Sakamoto, Y., and Ideno, Y. Improving Complex Nurse Care Activity Recognition Using Barometric Pressure Sensors. In **Human Activity and Behavior Analysis**, CRC Press, pp. 261–283, 2024. doi: 10.1201/9781003371540-18
- [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] Ni'ma Shoumi, M., and Inoue, S. Leveraging the Large Language Model for Activity Recognition: A Comprehensive Review. *International Journal of Activity and Behavior Computing*, pp. 11–12, 2024. doi: 10.60401/ijabc.21
- [12] Zhang, S., Li, S., Zhang, S., Shahabi, F., Xia, S., Deng, Y., and Alshurafa, N. Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on *Advances. Sensors*, 22(4), pp. 6–10, 2022. doi: 10.3390/s22041476
- [13] Jobanputra, C., Bavishi, J., and Doshi, N. Human Activity Recognition: A Survey. *Procedia Computer Science*, Vol. 155, pp. 698–703, 2019. doi: 10.1016/j.procs.2019.08.100
- [14] Alafif, T., Hadi, A., Allahyani, M., Alzahrani, B., Alhothali, A., Alotaibi, R., and Barnawi, A. Hybrid Classifiers for Spatio-Temporal Abnormal Behavior Detection, Tracking, and Recognition in Massive Hajj Crowds. *Electronics*, Vol. 12, No. 1165, pp. 2–3, 2023. <https://www.mdpi.com/2079-9292/12/5/1165>
- [15] Qian, H., Zhou, X., and Zheng, M. Abnormal Behavior Detection and Recognition Method Based on Improved ResNet Model. *Computers, Materials and Continua*, 65(3), pp. 2153–2167, 2020. <https://www.techscience.com/cmc/v65n3/38118>
- [16] Nassif, A. B., Darweesh, E., Al-Azzawi, A., and Jamous, R. Anomaly Detection in Human-Robot Collaboration Using LSTM Autoencoder. *Proceedings of the 2023 ACM/IEEE International Conference on*

- Human-Robot Interaction, pp. 334–336, 2023. doi: 10.1109/HRI.2023.1001234
- [17] Li, Z., Zhang, A., Han, F., Zhu, J., and Wang, Y. Worker Abnormal Behavior Recognition Based on Spatio-Temporal Graph Convolution and Attention Model. *Electronics*, Vol. 12, No. 2915, pp. 2–5, 2023. doi: 10.3390/electronics12132915
 - [18] Martín-Chinea, K., Ortega, J., Gómez-González, J. F., Pereda, E., Toledo, J., and Acosta, L. Effect of Time Windows in LSTM Networks for EEG-Based BCIs. *Cognitive Neurodynamics*, Vol. 17, pp. 385–398, 2023. doi: 10.1007/s11571-023-09895-2
 - [19] Dong, L., Fang, D., Wang, X., Wei, W., Scherer, R., and Woźniak, M. Prediction of Streamflow Based on Dynamic Sliding Window LSTM. *Water*, Vol. 12, pp. 6–11, 2020. doi: 10.3390/w12010006
 - [20] Ji, S., Zheng, X., and Wu, C. HARGPT: Are LLMs Zero-Shot Human Activity Recognizers? *Proceedings of the ACM International Conference on Human Activity Recognition*, pp. 1–3, 2024. URL: <https://example.com/hargpt-2024>
 - [21] Shoumi, N., and Wei, L. What Do Sensor-Based Human Activity Recognition Studies Tell Us About Multi-modal Integration? *Sensors*, 24(1), pp. 190–192, 2024. doi: 10.3390/s24010190
 - [22] Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M.: YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *arXiv preprint arXiv:2207.02696*, pp. 1–15, 2022. <http://arxiv.org/abs/2207.02696>
 - [23] Hochreiter, S., and Schmidhuber, J.: Long Short-Term Memory. *Neural Computation*, 9(8), pp. 1735–1740, 1997. doi: 10.1162/neco.1997.9.8.1735
 - [24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., and Jones, L.: Attention Is All You Need. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, pp. 1–15, 2017. <http://arxiv.org/abs/1706.03762>
 - [25] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., and Kaplan: Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*, pp. 6–7, 2020. <http://arxiv.org/abs/2005.14165>
 - [26] Yan, S., Xiong, Y., and Lin, D.: Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 6–7, 2018. <http://arxiv.org/abs/1801.07455>
 - [27] Zhou, K., Wu, T., Wang, C., Wang, J., and Li, C.: Skeleton-Based Abnormal Behavior Recognition Using Spatio-Temporal Convolution and Attention-Based LSTM. *Procedia Computer Science*, 174, pp. 424–432, 2020. doi: 10.1016/j.procs.2020.06.111
 - [28] Wei, X., and Wang, Z. TCN-attention-HAR: Human Activity Recognition Based on Attention Mechanism Time Convolutional Network. *Scientific Reports*, 14(7414), pp. 1–11, 2024. doi: 10.1038/s41598-024-64412-3
 - [29] Mori, S. Understanding Autism Through Cultural Models and Its Implications. *Doctoral Thesis*, Kobe University, Japan, pp. 118–122, 2021. doi: 10.18910/78901
 - [30] Fujioka, T., Garcia, C., and Inoue, S. (2025, March 17). Challenge: Abnormal Activity Detection in Individuals with Developmental Disabilities. *IEEE Dataport*. <https://dx.doi.org/10.21227/qfkw-sa40>
 - [31] Fujioka, T., Garcia, C., and Inoue, S. Toward Abnormal Activity Recognition of Developmentally Disabled Indi-

viduals Using Pose Estimation. *International Journal of Activity and Behavior Computing*, 2025(1), pp. 1–28. <https://doi.org/10.60401/ijabc.39>