

Title	九軸センサにより転倒検知
Author(s)	趙, 若辰
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19902
Rights	
Description	Supervisor: 藤波 努, 先端科学技術研究科, 修士 (情報科学)

修士論文

九軸センサーによる転倒検知方法の研究

ZHAO RUOCHEN

主指導教員 藤波努

北陸先端科学技術大学院大学
先端科学技術専攻
(情報科学)

令和07年3月

Abstract

With the aggravation of population aging, falling has become a serious social problem. In a series of falls, especially from the seat state, due to the concealment and potential danger of its movement characteristics, it has become a major challenge in health monitoring.

This study proposes an innovative fall detection framework that combines 9-axis sensors (accelerometer, gyroscope, magnetometer) and dynamic multimodal attention. The core of the system uses a model combining time convolution network (TCN) and dynamic multimodal multi-head attention mechanism, which can efficiently extract the characteristics of time series and adaptively adjust the weight according to the importance of each mode.

The experimental results show that the system is significantly better than the traditional methods in the accuracy, precision and recall of seat fall detection. In addition, through the interpretability analysis of the dynamic multimodal attention mechanism, the contribution of each mode in different fall scenarios was determined. This study provides a scalable and interpretable solution for multi scene real-time fall detection, which is of great significance to the development of intelligent elderly care and assistive technology.

目次

第 1 章 はじめに	1
第 2 章 転倒検出に関する研究	6
2.1 閾値による転倒検知	6
2.2 機械学習による転倒検知	7
2.3 Multi-source Signal Fusion	9
第 3 章 提案手法	12
3.1 モデル紹介	12
3.2 TCN (Temporal Convolutional Network)	13
3.2.1 因果畳み込みとダイレーテッド畳み込み	14
3.2.2 残差接続 (Residual connections)	18
3.3 注意機構(Attention Mechanism)	20
3.3.1 融合方法	20
3.3.1 マルチヘッド注意機構 Multi-Head Attention	21
第 4 章 実験・評価	23
4.1 実験	23
4.1.1 実験設備	23
4.1.2 実験方法	24

4.1.3 モデル設置	27
4.1.4 モデル訓練デバイス	28
4.2 評価.....	28
4.2.1 性能評価と比較	28
4.3 ブロック実験	29
4.3 ネットワークのバリエーション.....	31
第5章 おわりに	34
5.1 結果のまとめと従来方法との比較	34
5.2 重さの計算.....	36
5.2 結論.....	36
5.3 将来の仕事.....	37

図目次

図 3.1 : モデル構造	13
図 3.2 : dilated convolution 構造.....	13
図 3.3 : dilated convolution 構造 2	13
図 3.4 : Residual connections 構造	13
図 3.5 : 融合方法.....	21
図 4.1 : センサー設置方法	25
図 4.2 : 座位転倒.....	25
図 4.3 : 損失関数変化図	29
図 4.4 : 正確率変化図.....	30
図 4.5 : 精度と再現率変化図	30
図 4.6 : 損失関数変化図 2	31
図 4.7 : 正確率変化図 2.....	32
図 4.8 : 精度と再現率変化図	32
図 5.1 : MDATCN 混同行列.....	34
図 5.2 : 注意力重さ分布図.....	36

表目次

表 4.1: 加速度パラメーター	23
表 4.2: ジャイロスコープパラメータ	23
表 4.3: ADL の種類	23
表 4.4: 転倒の種類	23
表 4.5: 提案されたネットワーク構成	23
表 4.6: 性能評価方法	23
表 5.1: 関連研究正確率	35
表 5.2: 関連研究計算速度	35

第1章 はじめに

人間活動識別 (HAR) は、データ科学における急速に発展し、将来性のある分野の一つであり、ヘルスケアモニタリング、スマートホーム、転倒検出システムなど、さまざまな応用が広がっています。その中で、転倒検出は最近最も重要な研究焦点となっている。世界保健機関 (WHO) によると、転倒は世界で2番目に大きな事故死原因になっている [1]。このような行為がリアルタイムで監視または警報されれば、医療または救助時間が大幅に短縮され、重傷や死亡のリスクを最小限に抑えることができるようになる。そのため、高精度の転倒行動識別モデルを提案し、それを適切な場所と人の群れに適用することは重要な意義がある。従来の転倒検出技術は、センサータイプによって大きく3つの種類に分けることができる [2]：環境感知システム [3]、視覚に基づく転倒検出システム [4] とウェアラブルセンサー [5] である。

視覚ベースのセンサ (例えば監視カメラ) は、精度の高い特性があるため、現在市場で最も流行している転倒検出装置である、これは非侵入型の転倒検出方法である。Rougier らは、ビデオシーケンスを介して人の輪郭を追跡する形状マッチング技術を提案している [6]。この方法は、形状解析方法に基づいて輪郭から人体形状の変形を定量化する。最後に、ガウスハイブリッドモデルを使用して転倒と正常な活動を分類する。また、監視カメラの次に、深度カメラ (ToF カメラ) もこの分野で注目を集めている。(TOF) 深度カメラの最初の応用に関する研究は、2010年に Diraco らによって行われた [7]。彼らは、ランドマーク、キャリブレーションモード、またはユーザーの介入を必要としない視覚センサーに基づく新しい方法を提案した。しかし、ToF カメラは高価で画像解像度が低いという課題がある。その後、Kinect 深度カメラは2011年に Rougier らによって初めて使用された [8]。深さ情報から人体の重心高さと身体速度の2つの特徴を抽出する。簡単な閾値に基づくアルゴリズムを用いて転倒を検出し、全体的な成功率は98.7%であった。しかし、視覚センサーは弱い光に影響を受けやすく、さらに個人情報漏洩の問題も解決が難しいとされている。

環境感知システム環境センサーは、別の非侵入型転倒検出方法を提供する。アクティブ赤外線、RFID、圧力センサー、スマートタイル、磁気スイッチ、ドップラーレーダー、超音波センサー、マイクなどのセンサーにより、転倒に

よる環境変化を検出するために使用される。これは、受動的で普遍的な検出という革新的な方向性をこの分野に提供している[9]。超音波センサネットワークシステムは転倒検出システムの中で最も早い解決策の一つである。Horiらは、高齢者が住む空間に一連の空間分布のセンサーを置くことで転倒を検出できると考えており[10]、この検出方法では機器の固定設置場所の範囲や個人情報漏洩などしか検出できない問題に直面する必要があると考えている[9]。

ウェアラブルセンサーこれらは、電子センサーに基づくマイクロデバイスと定義することができ、保持者は、衣服の下、衣服の上、又は衣服の上に装着される[11]。これらの多くは加速度計を利用しており、一部はジャイロスコープなどの他のセンサーを統合し、ユーザーの位置情報などの情報を取得することを目指している。最近、マイクロ電気機械システム(MEMS)の発展に伴い、研究者は慣性センサーを小型ウェアラブルデバイスに統合して転倒検出を実現した。MEMSに基づく転倒検出方法は、導入コストが低く、環境適応性が高い上、ユーザーのプライバシー保護が可能であるため、転倒検出技術において人気のある研究分野となっている[9]。

複雑な場面や参加者のプライバシーを考慮する必要があるなど、さまざまな制限に起因するさまざまな要因を考慮して、本実験では、加速度計、ジャイロスコープ、磁力計を用いた9軸センサーで構成される慣性測定ユニット

(IMU)のみを使用する、3つの軸(x、y、z)の加速度を測定することによって線形運動と重力を検出する加速度計の9軸センサーのみを使用することにした。ジャイロスコープは、ローリング、ヨーイング、ピッチなどの回転速度を測定するために使用される。転倒に関する非常に詳細なデータを得ることができる。

転倒の方式も様々がある、転倒の影響と結果はさまざまな要因によって大きく異なる可能性がある。たとえば、歩行中、立っている時、寝ている時の転倒にはいくつか共通点があるものの、それぞれに顕著な違いがある。[9] El-Bendaryら(2013年)の研究では、転倒のタイプを「前方転倒」「側方転倒」「後方転倒」の3つに分類している[12]。しかし、座った状態での転倒という比較的特殊なカテゴリについては、ほとんど関心が払われていない。これは、座位での転倒における特徴の変化が非常に小さく、転倒の開始から終了までの時間も非常に短いためである。そこで、本研究の目標は座る転倒という問題の解決法を提案した。

転倒を検出するには、閾値技術と機械学習方法の2つの主要な方法がある。閾値の方法に基づくアプリケーションは容易に実現でき、計算作業量は最小である。転倒するかを検出できるが、再現率低い漏れる検査が多くという問題を解決できない。機械学習方法はより複雑で、もっと情報と数学知識を必要だが、検出率を高めることができる。機械学習と深層学習は過去10年間、画像認識、目標検出、音声認識、自然言語処理を含む多くの分野に革新的な変化をもたらした。典型的な行動認識問題として、多くの伝統的な機械学習と深層学習アルゴリズムはセンサーに基づく転倒検出問題を解決し、良好な効果を得た[9]。モデルには、サポートベクトルマシン (SVM)、ランダムフォレスト、深層ニューラルネットワーク (DNN)、畳み込みニューラルネットワーク (CNN)、長短期間記憶ネットワーク (LSTM)、再帰ニューラルネットワーク (RNN) が含まれる。しかし、加速度センサーやジャイロセンサーが取得したデータを用いて転倒検出に深度学習ネットワークを適用する場合、特に再帰アーキテクチャ (例えば単一 RNN モデル) を使用する場合、時間特性のグローバル情報を効率的かつ正確にキャプチャすることは困難であることがある。そこで予測パフォーマンスを向上させるために、本文は時間畳み込みネットワーク (TCN) という新しい方法を提案した。

転倒検知において、単一センサーいくつかの制約がある。例えば、センサーの故障によるデータ損失、空間的なカバー範囲の制限、測定精度や信頼性が単一センサーの性能に依存することがある。これらの問題を解決するために、データ融合技術が登場した。近年、急速に注目を集める研究分野となっている。複数センサーのデータストリームを統合することで、データ融合は信号対雑音比 (SNR) の大幅な向上、不確実性や曖昧さの低減、そしてシステムの堅牢性や耐障害性の強化を実現できる。また、冗長性を活用した多次元的な測定により、環境干渉に対する耐干渉性を高めるとともに、データの解像度や精度も向上させる。

特に、システム内でデータの事前処理を行うことで、アプリケーションロジックの複雑性を軽減し、開発プロセスを簡素化する。これにより、システムのモジュール化と保守性が向上し、スマートモニタリングシステムに対する効率的で柔軟なソリューションを提供する。[13] データ融合を実現するためには、最終決定における各モードの重要性を合理的に割り当てるために、異なるモードの特徴に重みを計算する必要がある。本研究では、IMU 信号 (加速度、角速

度、角度)を入力とし、異なる運動特徴の関係性を動的に捉えるために多頭注意メカニズム(MHA)を導入した。従来の静的な重み付け融合と比較して、本手法では、多頭注意を通じて異なる運動特徴間の相互関係を学習し、各モダリティ(加速度、角速度、角度)の重要性を状況に応じて適応的に調整する。例えば、転倒の瞬間においては角速度の特徴がより強調される傾向があり、転倒後の静止状態では角度の情報がより重要となる可能性がある。また、複数の注意ヘッドを利用することで、モデルが異なる視点から運動特徴を抽出し、小規模データセットにおける特徴学習能力を向上させるとともに、微弱な信号に対する感度を高めることができる。

そこで、まとめに言うと 本研究の貢献は以下になる。

1. 本研究では、座った状態から転倒するという特別な転倒を正確に検出するため、新しいデータセットを独自に構築した。このデータセットは、既存の転倒検出データセットでは十分にカバーされていない、座る動作に関連する細かい動きや特徴を捉えることを目的としている。具体的には、転倒に至るまでの一連の動作を記録することで、動作の遷移やパターンを明確化した。そして、我々も転倒以外の日常生活動作のデータも収集して、さらに、転倒に類似した日常生活動作(椅子に座る、地面に座る、物を拾う、歩く)のデータも同時に収集した。これにより、転倒動作と非転倒動作を正確に区別するための豊富な比較データを確保した。収集したデータはすべて手動でタグ付けを行い、転倒動作の発生タイミングやモード(加速度、角速度、など)を詳細にラベル化した。

2. 座位状態の転倒は、運動の振幅が小さく、特徴変化が微弱であり、発生時間が短いため、従来の転倒検出システムにとって大きな課題である。そのためモデル構造部分で、シグナル・フュージョン特徴融合に基づく動的マルチヘッドアテンションメカニズムと時間畳み込みネットワーク(TCN)を組み合わせた転倒検出システムの開発により、各特徴の重要性に応じて重みを自適応的に調整し、データの効果的な融合を実現する。同時に、TCNモデルを活用することで時系列特徴を効率的に捉え、転倒検出の精度と頑健性を向上させる。

3. 総じて言えば、本研究は、TCNと多信号融合(Multi-source Signal Fusion)を統合した転倒検出モデルを開発し、この手法を用いた研究はこれまでに報告されていないことを確認した。実験の結果、本手法の分類精度は98%

に達し、本研究がこの分野において有望な成果を示すものであることを証明した。

本文は以下のように構成されている。第二章では、九軸センサーに関連する研究、及び転倒検査に関する機械学習と深さ学習ネットワークの紹介、モデルの改善のための準備など、転倒検出関連する関連作業の開発と概要を説明する。第三章では、我々の MDATCN 全部の構成を紹介した。中には因果畳み込み (Causal Convolution) と拡張畳み込み (Dilated Convolution) と TCN ネットワークの関連コンポーネント、及び多モーダルとマルチヘッド注意機構の使用方法を述べる。第四章で、具体的な実験設備や実施方法などの関連情報及び実行する前処理プロセスを説明する。また、このセクションには、トレーニングプロセスや実験設定に関する詳細が追加されている。第五章で、実験の結果をディスカッションし、本研究の仕事をまとめて、将来の仕事を提案した。

第2章 転倒検出に関する研究

現在の段階では、9軸センサーと深層学習は転倒検知に関する研究に広く応用されている。ハードウェアの加速能力の向上とアルゴリズムの進歩により、研究者たちが高性能かつ高度なモデルを開発するための確固たる基盤が築かれた。本節では、転倒検知分野において利用される方法と9軸センサーに用いての転倒検知研究について説明する。

2.1 閾値による転倒検知

9軸センサーの特徴は、高集積度、低消費電力、高感度にあり、多種の複雑な応用範囲に適応できるようにしている。例えば、ウェアラブルデバイスでは、9軸センサーはフィットネス監視、活動追跡、転倒検出などの機能に広く使用されており、リアルタイムでユーザーの運動データを捕捉することで、その行動パターンを分析する。

Li, Qら2007年もジャイロと加速度計から派生した姿勢情報を使用し、正確かつ速い転倒検知を行う方法が提案された[17]。具体的には、ジャイロスコープと加速度計のデータを基に、人間の動静状態を推定し、転倒の有無を判断するものである。この方法により、転倒検知は91%の精度が達成された。この研究に基づいて、Yuanzhong Yらは9軸センサーを活用して、人間の姿勢変化を検出できるという特性を利用し、特に、人体姿勢を識別するため、もし上半身が地面と垂直方向と角度が常に 40° 以上であり、それが1秒以上続く場合、「横たわった姿勢（寝姿勢）」が検出されたと判断できることが発見した[18]。具体的には、実験の中で、1秒間のpitchまたはyawのすべてが 40° を超える場合、寝姿勢が検出される。そして、活動の強さにより、垂直方向の加速度が0.4g未満かつ2.6gを超える場合に転倒を検知する可能性が非常に高いことも示されている。クォータニオンとセンサーによる加速度を基に、重力方向の加速度値を計算することで、これらの特徴に基づく閾値判断を行った。この方法は従来の閾値法よりも良好な結果を得た。ただし、この研究では座位での転倒検知に関する考察が含まれておらず、また、座位転倒は非常に複雑動作なので、閾値決めることも難しいである。そして、機械学習と深層学習を比べて、閾値法特有の低いRecall率つまり転倒を漏れる問題が存在する。

2.2 機械学習による転倒検知

閾値の課題に対処するため、その後の研究では機械学習を活用した分類方法が試みられた。典型的な機械学習アルゴリズムとしては、SVM[19]、simplified decision-tree (SDT) [20]、k近傍法 (k-NN) [21] などが挙げられる。これらの方法はさまざまなシーンで使用され、優れた分類効果を得ている。しかし、機械学習アルゴリズムの成功は特徴選択に大きく依存しており、手で抽出された特徴が不十分である場合、転倒検知の精度に著しい影響を及ぼす可能性まだがある。座る転倒深層学習は自動的に特徴を抽出できるため、選択した。

機械学習ベースの転倒検知と比較して、深層学習アルゴリズムは特徴を自動的に選択できる能力を持ち、さらに強力な学習性能を備えている。現在、転倒検知分野で注目すべき性能を示しているいくつかの深層学習アルゴリズムが存在しており、特に CNN、RNN、LSTM は局所的な特徴を効果的に捉える能力を発揮している。1989年に提案されて以来、空間的特徴を入力とする CNN (畳み込みニューラルネットワーク) は広く応用され、転倒検出にも応用されている。多くの研究者が CNN モデルの再転倒検出における性能を向上させるための大量の作業を展開している。例えば、A. H. Fakhruddin らは一つはウェアラブルセンサーに基づいたの転倒検知システムを提案した[22]。最後の精度は 92.3%を達成した。A. H. Fakhruddin たちは時系列を画像に変換して CNN を認識させる。CNN に基づく時系列分類問題にもいくつかのメカニズムがあり、その中で応用が多いのは大体 2 種類である。最初のオプションは、時系列データを 1次元のメッシュ入力モデルとして直接処理することである。Zheng Y らは、多変数時系列を単変数時系列に分け、各単変数系列ごとに特徴学習を行う。第 2 のクラスは、1次元時系列データを 2次元画像特徴に変換し、その後モデルを入力して処理する。[23]例えば、多くの研究者がフーリエ変換やウェーブレット変換を用いて時系列データを周波数マップに変換する。画像を通じてより多くの関連特徴を得ることが望ましい。Lee S ら、これらの研究は本研究で参考としたが、組み込み式の計算要求を考慮すると、2次元は大量の計算資源を必要とし、1次元時系列で良い成績を得ることができれば、計算資源の浪費を回避することができるので、本研究は 1次元時系列入力を選択した。[24]もちろん、時系列を入力する RNN も広くに転倒検知に用いられている。E.Torti ら (2019) [25]もウェアラブルセンサーと RNN に基づいた

転倒検知方向を提出された。最後の正確率は 98%を達成した。その後、研究者たちも様々な RNN に基づいたの転倒検知方法を提出された。例えば、PerRNN[26]と IndRNN[27]などである。しかし、転倒検出タスクでは、モデルが長い時間範囲のコンテキスト情報を捕捉することができれば（すなわち、より長い時系列データを処理する）、通常、転倒は時間変化に関連する動的挙動であり、時間的な連続情報と結合して判断する必要があるため、検出の精度を高めることができる。しかし、これらの長時間シーケンスを処理する際に、逆伝播アルゴリズムを用いて訓練を行うと、勾配消失や勾配爆発の問題が発生する可能性があるこれらの問題を解決するために、LSTM という方法が提案された。ゲート機構を導入することで、長時間シーケンスにおける依存関係を学習できるようにする。LSTM も転倒検査で良い成績を収めた。Chuan-Bi Lin らも OpenPose スケルトンに戻つて、LSTM と普通 RNN 二つ方法を比べ、LSTM は正確率 95%といういい結果を得られた[28]。これらの研究は、転倒検出分野における深学習ネットワークモデルの潜在力を際立たせている。多くの検出アルゴリズムは高精度を証明し、センサーデータからユーザの活動状態を効果的に抽出した。しかし、より深い調査によると、単一アルゴリズムに依存するには限界があり、人間の転倒行為の異なる場面や変化に適応することが困難になることが明らかになった。対照的に、混合アルゴリズムは顕著な利点を示している。異なるアルゴリズムの利点を組み合わせることにより、これらの方法は、様々な環境と姿勢における転倒分類タスクの複雑さをよりよく処理することができる。Gang Yang らも CNN と LSTM 結合することにより、98.8%の正確率を得られた[29]。

前述のように、従来の深層学習アルゴリズム（RNN、CNN+LSTM など）では、高精度な分類が可能である一方で、今回の研究で利用すると以下のような課題がある。RNN 系の問題点は勾配消失・爆発により、長時間の時系列データを学習する際の性能が低下する。そして、計算が逐次的に行われるため推論速度が遅い。CNN 系の問題点は CNN だけでは時間的な依存関係を考慮できない。これらの問題を解決するため、近年、TCN (Temporal Convolutional Network) が提案され、時系列データ処理の分野で注目を集めている。TCN は並列処理でき、リアルタイムでの転倒検知にも適している。特に、組み込みシステムやモバイル環境においても、低遅延で動作可能である。

そして、本研究では、ウィンドウサイズが 100 サンプル（約 1 秒）に制限さ

れており、長期間の動作変化を監視することが困難である。単独のウィンドウのみを考慮すると転倒の長期的な変化を捉えるのが難しい。しかし、TCN の拡張畳み込み (dilated convolution) を活用し、複数の連続したウィンドウ間の関係を学習することで、転倒の長期的なパターンを捉える。

そこで、本研究の特徴提出した方法は TCN を選んだ。

2.3 Multi-source Signal Fusion

異なるモーダルを捕捉する各種センサーから大量のデータを収集し、ビデオ、無線探査と測距 (レーダー)、赤外線、光探査と測距 (Lidar)、慣性測定ユニット (IMU) などを含む。重要な課題の1つは、これらの異なるソースからの情報を統合するために効率的なデータ表現と構造を構築することである。しかしながら、これらのセンサーから得られる生データは、異種、複雑、不完全、および膨大である。多モーダルセンサー融合技術は、多モーダル情報を抽出し、組み合わせて、より信頼性の高い、より豊富な情報、より正確なデータを得ることによって、これらの課題に対応することができる。[30]

近年、多モーダル技術も様々の領域で利用されている。例えば、医学映像分野で画像分解、画像再構成、融合ルールなどのタモーダル融合方法を利用し、画像技術の物理的制約を取り除き、画像品質を向上させる[31]。自動運転分野で Prakash らも画像と LiDAR を注意機構で融合する新しい多モーダルトランスフォーマー「TransFuser」を提案した[32]。対話中の感情識別の分野にも、Chudasama, V. ら音声、視覚、テキストのモダリティを統合する新しい特徴抽出器とマルチヘッド注意機構を備えた多モーダル融合ネットワーク「M2FNet」を提案した[33]。

転倒検出の研究において、単一モダリティ (例: 加速度のみ) を用いた手法は、大規模なデータセットや高品質なデータを用いる場合には、比較的良好な分類性能を示すことがある。しかし、本研究で扱うような小規模データセットでは、いくつかの課題が発生する。

まず、本研究のような小規模データセットでは、深層学習モデルが転倒パターンの多様な特徴を十分に学習できず、一般化性能が低下する可能性がある。また、データのばらつきが不足しているため、単一モダリティでは限られた特徴量の範囲内で学習が行われ、モデルがトレーニングデータに過度に適合しやすく、新規データへの適応力が低下する。さらに、単一モダリティの手法には転倒パターンの識別能力に限界があり、例えば加速度データのみでは滑るよう

な転倒（前方転倒など）は比較的正確に検出できるものの、回転を伴う転倒（側面転倒や後方転倒など）の識別が困難になる。このような転倒パターンでは、角速度や姿勢情報など、他のモダリティの情報が不可欠となる。

したがって、本研究のような小規模データセットにおける転倒検出の精度向上のためには、複数の情報源を適切に統合し、特定のモダリティに依存しない手法が求められる。転倒検出の分野では、多モーダル融合技術も広く使用されており、Abroらは慣性センサー（IMU）と視覚データを統合した多モーダル転倒検出モデルを提案した[34]。この手法では、IMUデータをフィルタリングし、視覚データの特徴と統合した上で、多層パーセプトロン（MLP）分類器を用いて分類を行い、URFDデータセット上で88%の精度を達成した。

しかしながら、このような視覚情報を利用する手法にはいくつかの課題が残されている。例えば、視覚ベースの手法は遮蔽の影響を受けやすく、さらに個人情報リスクや高コストの問題が存在する。また、既存の融合方法は比較的単純な手法が多く、動的な特徴の変化を十分に考慮できていない。そこで、本研究では、これらの課題を解決し、低コストかつ高精度な転倒検出を実現するために、IMU（九軸センサー）に基づくSignal Fusion（多元信号融合）を提案する。IMUは加速度、角速度、角度情報を同時に取得できるという特性を持ち、単一モダリティの限界を克服し、より包括的な転倒検出を可能にする。

一方で、時系列データの特徴抽出においては、注意力メカニズム（Attention Mechanism）が重要な役割を果たすことが近年明らかになっている。注意力メカニズムは、自然言語処理（NLP）分野で大きな進展を遂げ、Transformer[14]やBERT[15]といった強力なモデルを生み出した。これらのモデルは、機械翻訳、感情分析、文書分類などのタスクで優れた性能を発揮しており、その成功を踏まえ、時系列データの処理にも応用されつつある。特に、転倒検出のような時系列データを扱うタスクにおいて、動的な特徴の変化を学習する上で注意力メカニズムの適用が期待されている。

しかしながら、現時点では、注意力メカニズムの時系列データへの応用はまだ十分に進んでおらず、[16]。そこで、本研究では、IMUベースの多元信号融合に注意力メカニズム（特にマルチヘッド注意機構）を組み合わせ、時間畳み込みネットワーク（TCN）と統合することで、座位状態における転倒検出の課題を解決した。本提案手法により、各モダリティの動的な重要性を考慮し、

異なる時間的特徴を適応的に学習することで、より高精度な転倒検出ができる。

第3章 提案手法

3.1 モデル紹介

まず、時間畳み込みネットワーク (TCN) とマルチヘッド (MHA) を組み合わせた動的注意力に基づく TCN モデル (MDATCN) の主な構造を簡単に紹介した。(図 3.1) この図は、加速度計、ジャイロスコープ、角度データを入力とするモデルのプロセスを示しています。それぞれのセンサーデータ (Accelerometer Input, Gyroscope Input, Angle Input) は、 x, y, z の三軸データとして入力され、各データは時間畳み込みネットワーク (TCN Block) を通じて処理される。この処理により、加速度 (Feature Vector 1)、角速度 (Feature Vector 2)、角度 (Feature Vector 3) それぞれの特徴ベクトルが生成されます。生成された特徴ベクトルは、多頭注意メカニズム (Multi-Head Attention) に入力される。このメカニズムは、3 つの特徴ベクトル (Query, Key, Value) を用いて特徴間の相関関係を学習し、各モダリティの重要性を動的に評価します。注意重み (Attention Weights) は、入力された特徴間の関係性を反映し、特定の運動状態においてどのモダリティが重要かを示します。

次に、これらの注意重みを用いて、重み付き和 (Weighted Sum) が計算され、加速度、角速度、角度の融合特徴が生成されます。その後、融合特徴に対して Dense レイヤーと Softmax を適用し、動的な重み付け (Dynamic Weights) が実現されます。これにより、各特徴ベクトルがその重要度に応じて適切に調整され、最終的に重み付き融合 (Weighted Fusion) により包括的な特徴ベクトルが得られる。

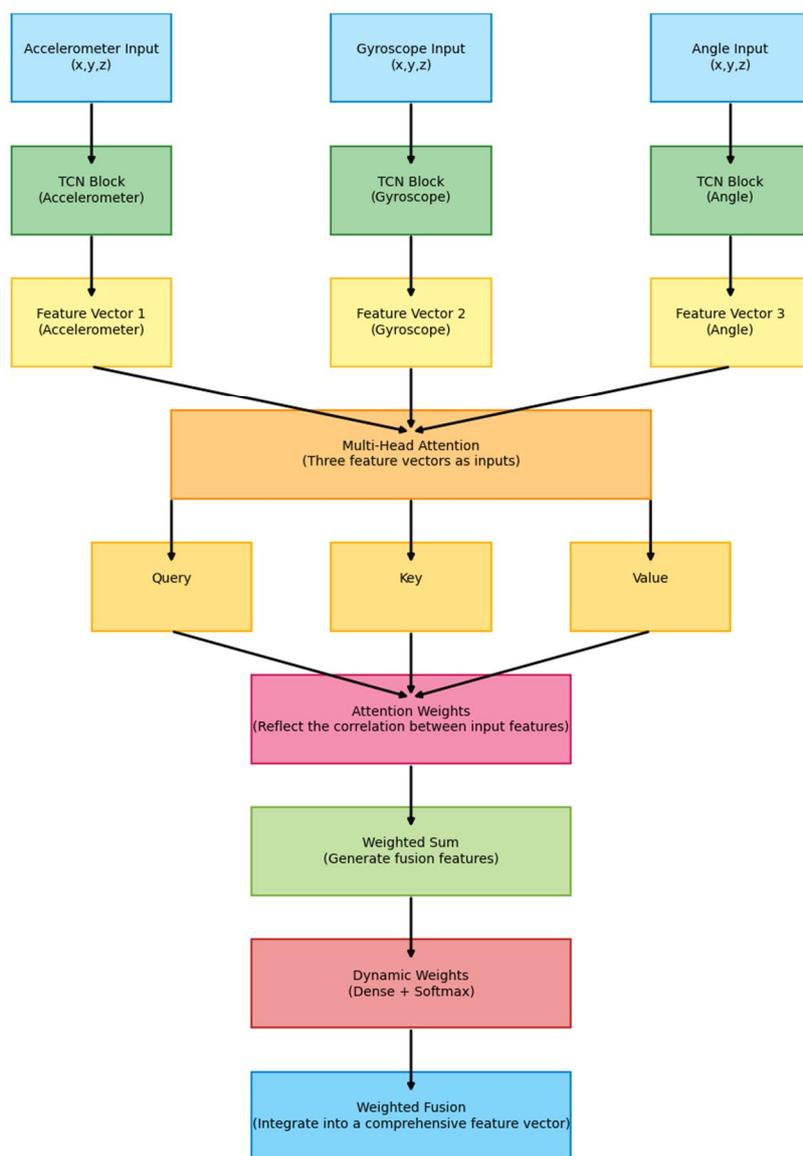


図 3.1 : モデル構造

3.2 TCN (Temporal Convolutional Network)

座位転倒の検出は、高齢者の安全確保において重要な課題であるが、その実現にはいくつかの困難が伴う。座位転倒は動作の振幅が小さく、センサー信号の変化も微弱であるため、従来の手法では特徴を十分に抽出できない。また、転倒が短時間で発生するため、時系列情報を適切に捉えることが難しく、既存のモデルでは時間的依存関係の学習が不十分となる可能性がある。さらに、座位転倒は日常的な動作（例：座る、体を傾ける、拾い物をするなど）と類似しており、これらを誤検出するリスクが高い。

これらの課題を踏まえ、本研究では 時間畳み込みネットワーク (TCN) を採用した。TCN は、カジュアル畳み込み (Causal Convolution) と拡張畳み込み (Dilation Convolution) を組み合わせることで、長時間の依存関係を効率的に学習できる。これにより、短時間で発生する座位転倒の前後の時系列パターンをより適切に捉えることが可能となる。また、TCN は畳み込み構造に基づいており、並列計算が可能であるため、従来のリカレントネットワークと比較して学習速度が速く、計算コストも低い。

さらに、TCN は多層の畳み込みにより、微細なセンサー信号の変化を効果的に捉えることができる。この特性により、座位転倒の特徴をより明確に抽出し、日常的な動作との識別精度を向上させることが期待できる。従来の手法では、類似動作との誤検出が課題となっていたが、TCN の時間的な特徴抽出能力によって、この問題を軽減できる。

これらの理由から、本研究では座位転倒の検出に TCN を用いることを決定した。次のセクションでは、具体的な TCN の構造とその特性について詳しく説明する。

3.2.1 因果畳み込みとダイレーテッド畳み込み

本研究では、入力データは 9 軸センサー (加速度計、ジャイロスコープ、角度センサー) によって固定サンプリングレートで収集された時系列信号である。そこで、膨張の因果畳み込みネットワークを紹介する前に、シーケンスモデリングタスクの基本的な性質を説明する必要がある。センサーで取得した時系列データ x_0, x_1, \dots, x_t は入力信号であり、アナログ変換を経てモデルに入力して、対応する出力、すなわち転倒検出の分類結果を予測したい。このプロセスでは、重要な制約は、以前に観察された入力データを使用してターゲット出力を予測し、将来の情報の漏洩を回避することを必要がある。

数学的形式から見ると、シーケンスモデリングタスクの目標は、予測結果と実際の結果との間の損失値を最小化するマッピング関数 $f(\cdot)$ を見つけることである。関数 $f(\cdot)$ は、次の式で表すことができる。

$$y_t = f(x_0, x_1, \dots, x_t)$$

この中で、 y_t は時間 T の予測出力、 x_0, x_1, \dots, x_t は前の入力信号だから、シーケンスモデリングに適したモデルを設計するには、以下の条件を満たす必要がある。まず、長さが異なるシーケンスを扱えること、次に長期的な関係を捉えられること、そして時間の順序を正しく保持することである。また、シ

一ケンス全体でパラメーターを共有する仕組みも重要である。これらの条件を満たす RNN (リカレントニューラルネットワーク) は広く使われているが、時間畳み込みネットワーク (TCN) が RNN を上回る性能を示すことがある。[35]TCN には次の特徴がある。まず、TCN は因果畳み込み層を使用しているため、未来の情報の過去への漏洩を回避でき、それによって時系列予測タスクの因果性を保証することができる。次には、シーケンス長が長いタスクを処理する場合、十分な歴史情報を得るために、ネットワーク深度または畳み込みコアサイズは TCN 入力シーケンス長が長くなるにつれて増加する。その結果、勾配爆発や勾配消失の問題が発生する。そこで、この主な欠点を解決するために dilated convolution を導入した。dilated convolution は、畳み込み核間に間隔を加えることにより、感受野を効果的に拡張した。図 3.2 により、この図は、畳み込み処理における三種類の異なる手法 (通常の畳み込みおよび膨張畳み込み) の接続パターンと、それに伴う受容野の拡張能力を示したものである。それぞれの手法が、時間系列データのモデリングにおいて持つ特徴や適用可能性を視覚的に説明している左側の図は、ストライドが 2 (stride=2)、カーネルサイズが 3 (kernel size=3) の畳み込み処理を表している。この手法では、ストライドにより入力ノードを間引きながら処理を行うため、計算量を削減できるという利点がある。しかしながら、入力データの一部が無視される結果、時間系列データにおける連続的な情報を捉える能力が低下する。したがって、この手法は計算効率を重視するタスクに適している一方で、高い精度が求められるタスクには不向きである。中央の図は、ストライドが 1 (stride=1)、カーネルサイズが 3 (kernel size=3) の通常の畳み込み処理を示している。この方法では、全ての入力ノードを逐次処理するため、時間系列データの細やかな変化を詳細に捉えることが可能である。しかしながら、受容野の拡張速度が遅いため、長期間にわたる依存関係を学習する際には限界が生じる。短期間における特徴抽出には適しているが、長期的な依存性が重要となるタスクにおいては効率性の課題が残る。右側の図は、ストライドが 1 (stride=1)、カーネルサイズが 3 (kernel size=3)、膨張率 (dilation rate) が 2 の膨張畳み込み (dilated convolution) を表している。この手法では、カーネル間に間隔を設けることで、計算量を増加させることなく受容野を効率的に拡張することが可能である。膨張畳み込みの最大の特長は、短期的な情報を保持しながら長期間の依存関係を捉える能力

を向上させる点にある。このような特性は、転倒検出のように、時間的変化や長期間にわたる依存関係を重視するタスクにおいて非常に有効である。

時間系列データのモデリングにおいて、膨張畳み込みは優れた性能を発揮する。通常の畳み込みが持つ高い解像度の特徴と、ストライド畳み込みの計算効率を兼ね備えつつ、指数関数的に受容野を拡張できるためである。このような特性により、膨張畳み込みは短期的な情報と長期的な依存関係の双方を効率的に処理可能であり、特に時間的な特徴を重視する応用分野において強力な手法となっている。

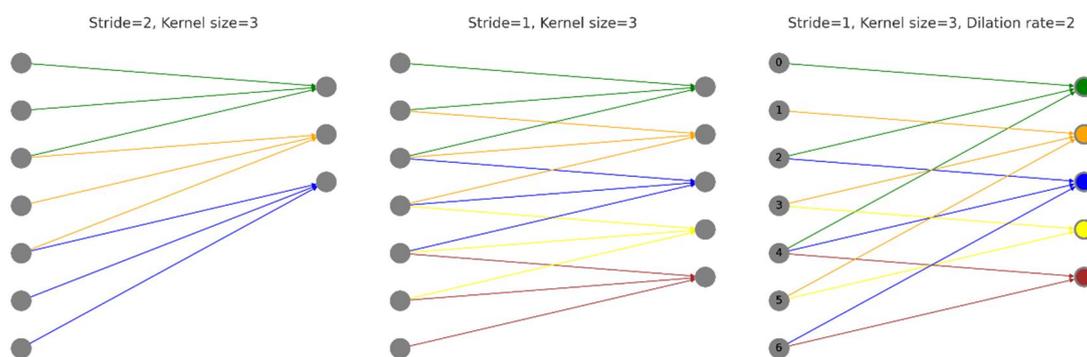


図 3.2 : dilated convolution 構造

例えば、図 3.3 で膨張因果畳み込みの構造は図 3.3 に示すように、2層畳み込み、膨張因子を含む d はそれぞれ 1、2、畳み込みコアサイズ k は 3 である。この構成では、出力の y_t は受容野 $X_{\{t-4\}}$ に達することができる。

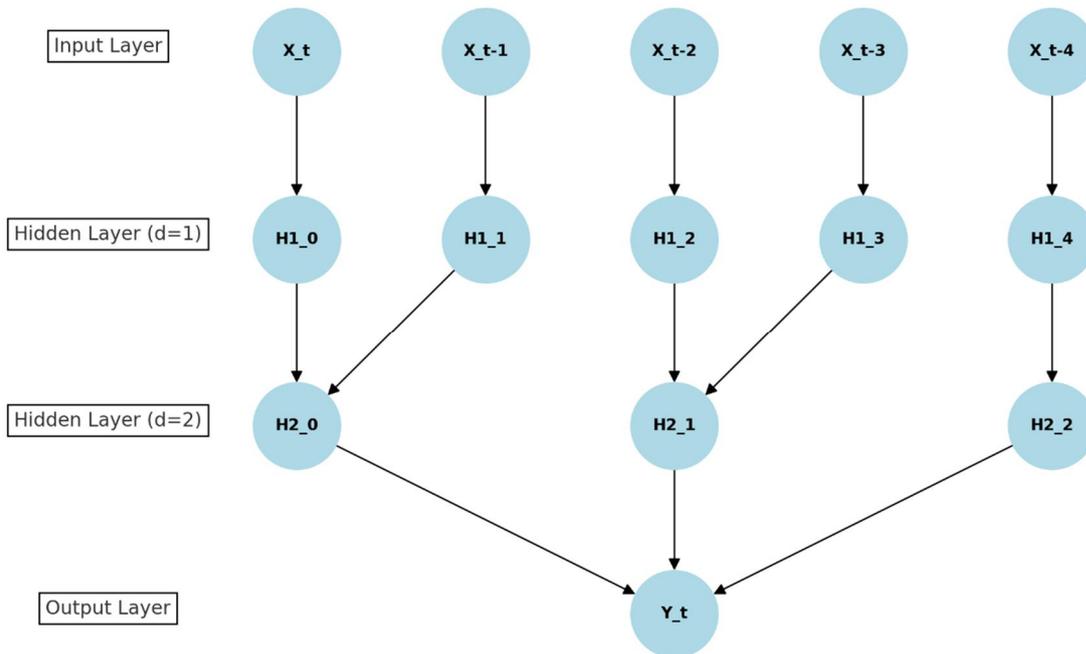


図 3.3 : dilated convolution 構造 2

次に、本研究で使用される膨張因果畳み込み (Dilated Causal Convolution) の具体的な設計について説明する。まず、膨張因果畳み込みの受容野

(Receptive Field, R_N) が入力系列全体の長さ (L_s) を完全にカバーする必要がある。この条件を満たすことで、時系列全体の情報を利用した正確な出力が得られる。

膨張因果畳み込みを適切に設計するには、以下の条件が必要となる：

$$R_N \geq L_s$$

ここでは R_N 総感受野であり、ネットワークのすべての畳み込み層を通じて、出力ノードが参照可能な入力系列の範囲を指す。 L_s は入力シーケンスの長さである。ライドスライス処理を行うことで、全てのデータが一定の時間ステップ (time steps) に固定されている。膨張因果畳み込みの総感受野を入力系列全体に広げるには、適切な畳み込み層数 (N) が必要である。この層数は以下の式で求められる：

$$N \geq \log_2 \left(\frac{L_s - 1}{k_s - 1} + 1 \right)$$

本研究では $L_s = 100$ $k_s = 3$ (畳み込みカーネルサイズ) そして、結果により 6 層膨張因果畳み込み層を使用することで、モデルの総感受野が入力シーケンス全体 (100 ステップ) をカバーできる。しかし、実際の実験において 6 層構成

のモデルでは計算コストが増加し、収束速度が遅くなるだけでなく、過剰な感受野によって短期的な特徴が埋もれ、転倒検出精度が低下する傾向が見られた。一方、層数を2層に減らしたモデルでは、短期的な特徴がより効果的に抽出され、精度が向上したと同時に、訓練時間の短縮も達成された。これは、座位転倒のような短時間で特徴的な動作が発生するタスクにおいて、長期的な依存関係を過剰に考慮する必要がなく、簡素化されたモデルの方が汎化性能や実用性が高まることを示している。以上の結果から、本研究では2層構成の膨張因果畳み込みが最適であると結論付けられた。

3.2.2 残差接続 (Residual connections)

残差接続の核心は、各層が入力に対する「残差 (Residual)」を学習することであり、直接出力を学習するのではない。この仕組みは、各層の出力と入力の間スキップ接続 (skip connection) を追加することで実現される。例えば、ある層の入力が x だとすると、目標は $H(x)$ というマッピングを学習することである。従来のネットワークでは $H(x)$ を直接学習するが、残差接続ではショートカットを導入することにより、ネットワークが実際に学習するのは以下の式：

$$F(x) = H(x) - x$$

したがって、目的関数は次のようになる：

$$H(x) = F(x) + x$$

ここで、 $F(x)$ はこの層のネットワークにおけるパラメータ化された部分であり、 x は層を直接スキップする入力である。

図 3.4 は、残差接続がどのように TCN モデルに応用されているかを具体的に示している。この図では、入力信号 (緑色のボックス) が畳み込み層 (Conv1D) を通過して処理される一方、赤い矢印で示されたスキップ接続によって入力信号がそのまま後続の層に直接伝達されている。スキップ接続は、入力特徴を保持しながら、ネットワークの出力と統合する役割を果たしており、これにより重要な特徴が損なわれることなく、学習効率を高めている。

本研究では、この残差接続を導入した時系列畳み込みネットワーク (TCN) を設計し、より効率的な深層学習モデルを構築した。スキップ接続を通じて入力信号を直接後続の層に加算することで、原始的な入力特徴を保持しつつ、高次の特徴を融合する設計を実現している。また、各畳み込み層の出力には層正規化 (Layer Normalization) を適用し、特徴分布をバランス化し、学習の安定性を

確保している。このような残差接続の活用により、ネットワークの学習効率が向上するとともに、モデルの頑健性や汎化性能が大幅に向上している。

特に、本研究が焦点を当てている座位転倒検知タスクでは、短時間の動作特徴を正確に捉えることが重要である。座位からの転倒は、立位の転倒に比べて動作の継続時間が非常に短く、特徴的な変化が瞬間的に発生する。このような短時間の特徴を見逃さないためには、時間依存性を適切に捉えるネットワーク設計が求められる。また、座位からの転倒は、物を拾う動作や座り直しといった日常動作と非常に類似しており、それらを区別することも課題の一つである。

スキップ接続を組み込んだ本モデルは、これらの課題に対応するために、入力信号を保持しながら残差を効率的に学習し、短時間内の微細な特徴を捉える能力を強化している。これにより、座位転倒検知の精度を大幅に向上させ、誤検知を抑えることを可能にしている。

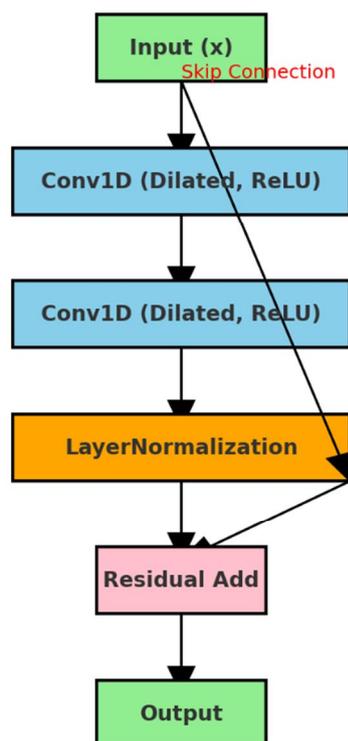


図 3.4 : Residual connections 構造

3.3 注意機構(Attention Mechanism)

3.3.1 融合方法

融合には、データ融合、特徴融合、決定融合などの様々な方法も含まれている。データ融合は、センサーレベルにおいて生データを直接統合する手法である。特徴融合は、各センサーから抽出された特徴を統合する手法であり、融合のタイミングによって前期融合 (early fusion) と後期融合 (late fusion) に分類される。前期融合は、異なるモダリティの特徴を統合した単一の特徴ベクトルを作成し、その後の処理を行う方法である。後期融合は、各モダリティの特徴を独立に処理した後、高次層において統合を行う手法である。決定融合は、各モダリティを独立に処理し、それぞれの予測結果を融合する手法である。研究では特徴融合における後期融合を使用する予定である。図 3.5 のように深層層学習により、加速度角速度角度の特徴を取った後融合する。本研究では、特徴融合の中でも後期融合を採用する。座位転倒検知においては、加速度、角速度、角度といった異なるモダリティのデータを扱うが、それぞれのデータは異なる物理的特性を持ち、直接的なデータ融合を行うと冗長な情報やノイズの影響を受けやすくなる。そのため、各モダリティの特徴を個別に学習し、それらを適切に統合することが求められる。後期融合では、各モダリティに特徴抽出ネットワークを用いることで、低層におけるモダリティ間の干渉を抑えつつ、情報を深く抽出することが可能である。

また、後期融合では注意機構を活用することで、各モダリティの重要性に応じた重み付けが可能となり、より適応的な情報統合が実現できる。特に、加速度は高速な動作、角速度は回転運動、角度は姿勢変化の検出に有効であり、それぞれの特徴を適切に捉えた後で統合することで、座位転倒のような微細な動作の検出精度を向上させることができる。

以上の理由から、後期融合は情報統合の効率性と柔軟性に優れ、本研究の目的に最適な手法である

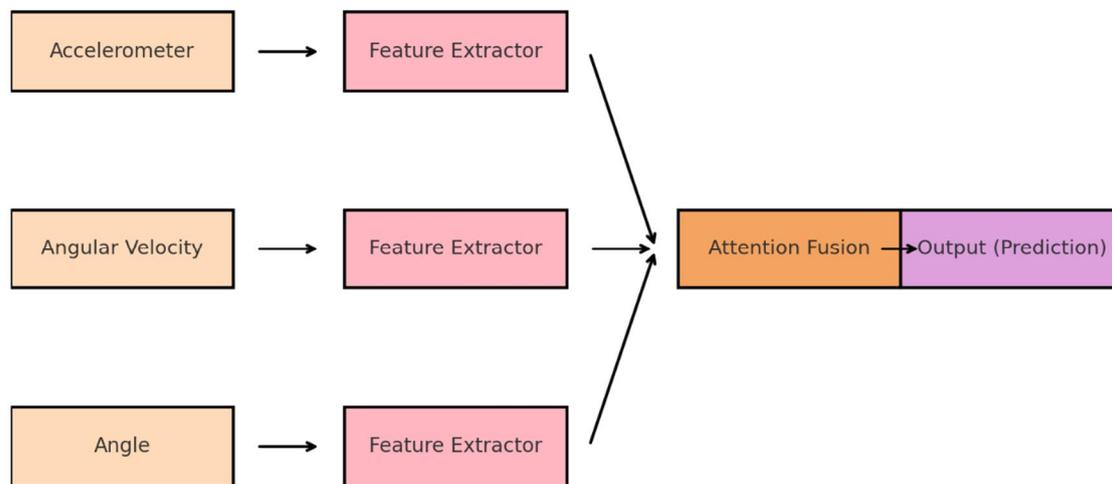


図 3.5 : 融合方法

3.3.1 マルチヘッド注意機構 Multi-Head Attention

本研究では、加速度、角速度、角度の特徴を統合し、信号融合を最適化するために マルチヘッド注意機構 (Multi-Head Attention, MHA) を導入した。MHA は Transformer アーキテクチャの中核を成す機構であり、シーケンス内の要素間の関係を捉えることに優れている。特に、本研究の目的である 座位での転倒 (Seated Fall) 検出では、異なるモダリティ間の相互作用を考慮することが不可欠であるため、MHA を活用することで信号の統合的な表現を獲得することを目指した。MHA の核心的な考え方は、入力特徴を複数の異なる空間に投影し、それぞれ独立した注意機構を並列に適用することである。これにより、単一の注意機構 (シングルヘッド注意) では捉えられない、異なる視点の特徴を同時に学習することが可能となる。本研究では、8 個のヘッド (head) を用いたマルチヘッド注意を採用し、各ヘッドが異なる特徴表現を学習できるように設計した。本研究では、時間畳み込みネットワーク (TCN: Temporal Convolutional Network) を用いて、加速度、角速度、角度の特徴をそれぞれ抽出した。得られた特徴を クエリ (Q)、キー (K)、バリュー (V) としてマルチヘッド注意機構に入力し、各モダリティ間の相互関係を学習させた。

$$Q, K, V = \{Features\}$$

次に、マルチヘッド層の中で、 Q, K, V に線形変換を行い、複数のヘッドの注意力表示を生成する。公式は：

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

次に、複数ヘッドの結果が Concat で結合された後、入力フィーチャー空間にマッピングされる。公式は：

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

そして、Lambda 層を用いて時間次元の平均値を求め、各モードの全体的な注意表現を得た。公式は：

$$\{attention_pooled\} = \frac{1}{T} \sum_{\{t=1\}}^{\{T\}} \{Attention\}(Q, K, V)$$

Dense レイヤを使用してプール化されたフィーチャーに動的重みを生成し、softmax を経て正規化させる。

$$\{attention_weights\} = \{softmax\}(W_{\{dense\}} \cdot \{attention_pooled\})$$

最後、動的重みとモダリティフィーチャーを Multiply で1つずつ乗算し、すべての加重後のモダリティ特徴を concatenate を使って結合する。

この設計により、MHA を活用した動的重み付けによって、各モダリティの相対的な重要度を適応的に調整し、高精度な座位での転倒検出を実現することが可能となった。

第4章 実験・評価

4.1 実験

4.1.1 実験設備

本研究は、知識科学倫理審査会議（承認コード：KSEC-A2024121302）の承認を得て実施した。装置の面では、Witmotion の BWT901CL 9 軸センサーを選択した。この製品は MEMS 技術に基づく高性能運動姿勢測定システムである。3 軸ジャイロスコープ、3 軸加速度計 3 軸電子コンパスなどのモーションセンサーを含む。各種高性能センサーを集積し、高動的カルマンフィルタリングを結合することにより、高精度で高動的リアルタイム補償の 3 軸姿勢と角度を直接提供することができる。出力データレート 0.2-200 Hz は調整可能で、同時に Bluetooth 2.0 と Typec シリアルポートをサポートし、シリアルレートは 115200 bps である。具体的なパラメータデータを下表に示す[36].

パラメーター	条件	典型値
レンジ		±16g
分解能		0.0005(g/LSB)
RMS ノイズ	帯域幅=100Hz	0.75~1mg-rms
静止ゼロドリフト	水平放置	±20~40mg
帯域幅		5~256Hz

表 4.1: 加速度パラメーター

パラメーター	条件	典型値
レンジ		±2000°/s
分解能	±2000°/s	0.061(°/s)/(LSB)
RMS ノイズ	帯域幅=100Hz	0.028~0.07(°/s)-rms
静止ゼロドリフト	水平放置	±0.5~1°/s
帯域幅		5~256Hz
ゼロバイアス安定性		≤10deg/h
感度		≤0.015deg/s rms

表 4.2: ジャイロスコープパラメータ

センサーは加速度角速度角度の 3 軸データを収集した後、Bluetooth を通じて直接伝送し、計算資源及び転倒発生時間も 1 s 程度であることを考慮してサンプリング率を 100 HZ に設定し、上位機は surface pro 8 を用いてデータ受信保存と整理を行う。具体的なデバイスインフォメーション Cpu : 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60 GHz (最大動作周波数 2.61 GHz)
RAM : 8.00 GB (使用可能容量 : 7.84 GB)

4.1.2 実験方法

A 設置方法

ウェアラブル転倒検出器は、実際の転倒イベント、特に実際の転倒と転倒に似たアクティビティを迅速に検出する必要がある。より効果的な転倒データセットと日常生活動作データセットを得るためには、人体の転倒活動にどの位置が反応しやすいのかという問題と、いくつかのセンサーが必要なのかという問題がある。センサーの数を増やすことで、取得できるデータの量や多様性が向上し、検出精度の向上が期待できる。しかし、センサーの増設は、ハードウェアコストの上昇に加え、システムの実用性や運用上のメンテナンス負担を増大させる要因ともなる。また、多数のセンサーを用いたシステムでは、センサーフュージョンの複雑化により、システムの信頼性低下やユーザーの装着負担の増加が問題となる。Khalifa, S らの研究 [37] では、センサー数を削減することで、一定の検出精度の低下はあるものの、デバイスのエネルギー効率が向上し、実用性の観点で有利であることが示されている。これらの点を考慮し、本研究では、シンプルで実用的なシステム構成を維持するために、単一の 9 軸ウェアラブルセンサーを採用した。センサーの数を決めるあと、センサーの設置位置を決める必要があり、多くの研究者が転倒事件を効果的に反映できる体の位置を見つけるために大量の仕事をしている。Maarit Kangas らの研究 [38] によると、センサーを腰に置くことで転倒の条件により効果的に反応し、より完全な身体運動データを得ることができることが分かった。だからセンサーは最後に腰の位置に置くことを決めて、図のように設置された。

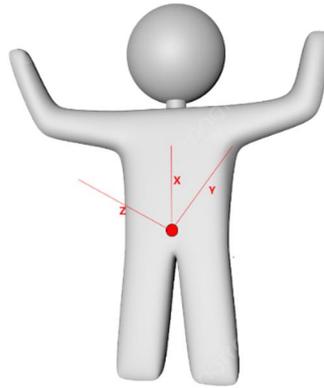


図 4.1：センサー設置方法

B シミュレーターされた転倒と日常生活動作（ADL）の研究

転倒負傷のリスクと安全性を考慮した本実験には、20～30歳前後の青年被験者のみを募集し、合計17人を募集し、それぞれ男性8人と女性9人、体重範囲は46 kg～104 kg、身長は155 cm～186 cmだった。実際の転倒状況をシミュレートするのは難しいので、安全な環境で転倒をシミュレーションすることで、3つの転倒姿勢を選択した。それぞれのパターンについて10回ずつ繰り返しデータを収集した。



図 4.2：座位転倒

座位転倒に焦点を当てるため、選定した転倒パターンは以下の通りである：椅子から滑り落ちる前方転倒、椅子から横に傾いて倒れる側方転倒、および立ち姿から後ろに倒れる後方転倒。転倒研究以外にも日常生活の動作データ収集を行い、転倒と似た動作を選択した。選定した動作は、日常生活で頻繁に行われる基本的な動作に基づいています。具体的には、「椅子への着席」「地面に座る動作」「通常の歩行」「物を拾う動作」の4つである。最初の動作は、立った状態から椅子に座る「椅子への着席」である。この動作は短時間で完了し、脚から腰へと重心が移動する基本的な動作である。次に、「地面に座る動作」は、立った状態から地面に座る動作であり、椅子に座る動作よりもさらに身体を屈曲させる必要がある点が特徴である。この動作では、重心がより低い位置に移動する。また、「通常の歩行」は、一般的な速度で8sぐらい歩く動作を指す。最後に、「物を拾う動作」は、立った状態から身体を前に傾けて地面にある物を拾い上げる動作である。この動作では、腰を屈めると同時に重心が前方に移動することが特徴である。そこで模擬転倒と日常生活動作を合わせて7種類の活動を行い、次の表の通り：

N	Types of ADL	Is fall-like
1	Sitting	Yes
2	Sitting down	Yes
3	Walking	No
4	Picking up	Yes

表 4.3:ADL の種類

N	Types of FALL
1	seated_front_fall
2	seated_side_fall
3	vertical_fall

表 4.4: 転倒の種類

C 数据处理

BWT 901 CL の 9 軸センサーには、カルマンフィルタリングによるノイズ削減機能が搭載されており、これによりデータノイズ除去のプロセスを省略することが可能であった。また、ジャイロスコープの性能が極めて優秀であるため、ドリフトの問題についてもほとんど処理を行う必要がなかった。収集したデータは、転倒および未転倒の二つのクラスに分類し、それぞれにラベルを付与した。深層学習を円滑に進めるため、まずデータクリーニングを実施し、空白データや無効データを削除した。その後、スライディングウィンドウ処理を適用し、ウィンドウサイズを 100 タイムステップ、スライドサイズを 10 ス

トップに設定した。ウィンドウサイズ（100 タイムステップ）は、センサーのサンプリングレート（100 Hz）と転倒動作の平均持続時間（約 1 秒）に基づき設定した。この設定により、転倒動作を完全にカバーするデータが確保できるスライドサイズ（10 ステップ）は、データの連続性を保ちながら計算コストを抑えるために選択した。0.1 秒のオーバーラップを設けることで、動作の特徴がウィンドウ間で失われるリスクを軽減している。さらに、スライディングウィンドウ処理の際には、転倒データと未転倒データが混在しないよう留意した。具体的には、各動作の終了時点でウィンドウを切り離し、100 ステップに満たないデータについてはゼロ埋めを行い統一した。このような手法により、すべての動作データが正確かつ一貫性をもって整理され、モデルの学習に適した形式で提供されるようになった。

4.1.3 モデル設置

ネットワーク構造のモデル設定について、各層のパラメータは表 4.5 のように示されている。本ネットワークの入力データは 9 軸センサーから得られる多モーダルの時系列データであり、まず 5 つの積み重ねられた TCN モジュールを通じて時系列特徴の抽出が行われる。TCN モジュールでは膨張畳み込みを使用して受容野を拡大し、長期間の依存関係を効果的に捉えることができる。その後、特徴はクロスモダリティ注意機構を通じて、異なるモダリティ間の情報相互作用をモデル化する。ネットワークは、動的重み生成モジュールを通じて各モダリティに適応的な重みを割り当て、これらの重みとモダリティ特徴を組み合わせた後、特徴融合層を介して全体的な特徴を統合します。最後に、全結合層で特徴をさらに処理し、出力層で分類を完了する。

レイヤー	パラメータ
1. 入力層	加速度データ、角速度データ、角度データの 3 つのモダリティ入力を受け取る。
2 TCN block1	TCN Block 1: Kernel size = 3, Filters = 64, Dilation rate = 1。感受野の構築を開始。
3 TCN block2	TCN Block 2: Kernel size = 3, Filters = 64, Dilation rate = 2。
7. クロスモダリティ注意機構	Multi-head Attention: Heads = 8, Key dimension = Feature size, Mean pooling。

8. 動的重み生成モジュール	Dense Layer: ソフトマックス活性化関数 (Softmax activation) を使用し、3つの重みを生成
9. モダリティ特徴融合モジュール	重み付けされた特徴量を融合し、各モダリティの重要性を考慮した特徴量を生成。
10. 全結合層	Dense Layer: Neurons= 128, Activation function= ReLU,
11. 出力層	Dense Layer: Neurons= 1, Activation function)= Sigmoid, 二値分類タスク用

表 4.5:提案されたネットワーク構成

4.1.4 モデル訓練デバイス

私の深層学習モデルは、TensorFlow (Keras) を使用して構築され、Scikit-learn による前処理と評価が行われている。計算プラットフォームには、Intel i9-12900K プロセッサ (クロック周波数 3.70 GHz)、32 GB メモリ、および 8 GB の NVIDIA GeForce RTX 3080 グラフィックスカードが搭載されている。すべてのモデル Glorot Uniform 初期化を使用し、Adam オプティマイザーを用いて逆伝播によるトレーニングを行った。バッチサイズは 32、トレーニングエポック数は 40、初期学習率はデフォルト値の $1e-4$ に設定されている。

4.2 評価

4.2.1 性能評価と比較

本研究は訓練中で二値交差エントロピー (Binary Crossentropy) を損失関数として使用される。さらに、モデルの性能を評価するために混同行列を使用した。モデルのテスト時には、正確率、精度と再現率 (リコール) 及び F1 点数もモデル評価の指標として導入した。

- (1) $Accuracy = (TP + TN)/(TP + TN + FP + FN)$
- (2) $Precision = TP/(TP + FP)$
- (3) $Recall = TP/(TP + FN)$
- (4) $F1-Score = 2 * Precision * Recall / (Precision + Recall)$

Predict	Positive	Negative
True	True Positive (TP)	True Negative (TN)
False	False Positive (FP)	False Negative (FN)

表 4.6:性能評価方法

4.3 ブロック実験

本節では、TCN ブロックの深さを調整することでネットワークを最適化しようとした。前述のように、モデル構築段階において、TCN ブロック内の残差

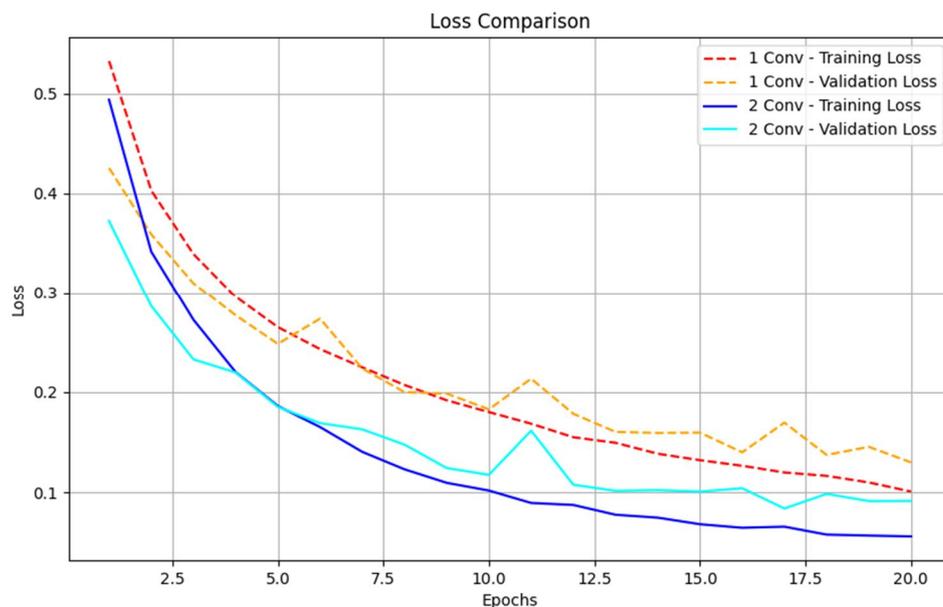


図 4.3：損失関数変化図

マトリックスが2層の場合にTCNブロックの性能が最適であると仮定した。この仮説を検証するために、まずTCNブロック内の畳み込み層を1層に減少し、比較実験を行った。その結果、残差マトリックスが2層のTCNブロックでは、4つの評価指標（損失関数、再現率、精度、適合率）がトレーニング過程で大きく変動するのに対し、1層の残差マトリックスを持つTCNブロックでは変動が少ないことが分かった（図10参照）。しかし、結果により2層のTCNブロックでは、すごくいい成績を取った。そのため、効率性と精度まとめて、そして、転倒検測はという命と関連する技術なので、考えた後、TCNブロック内の残差マトリックスを2層にすることを決定し、より良く結果を目指す

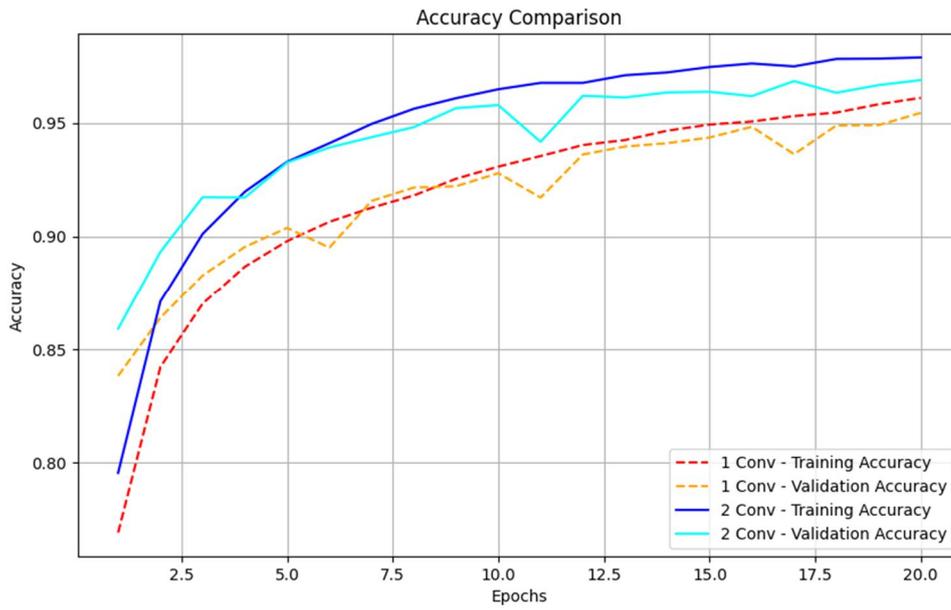


図 4.4 : 正確率変化図

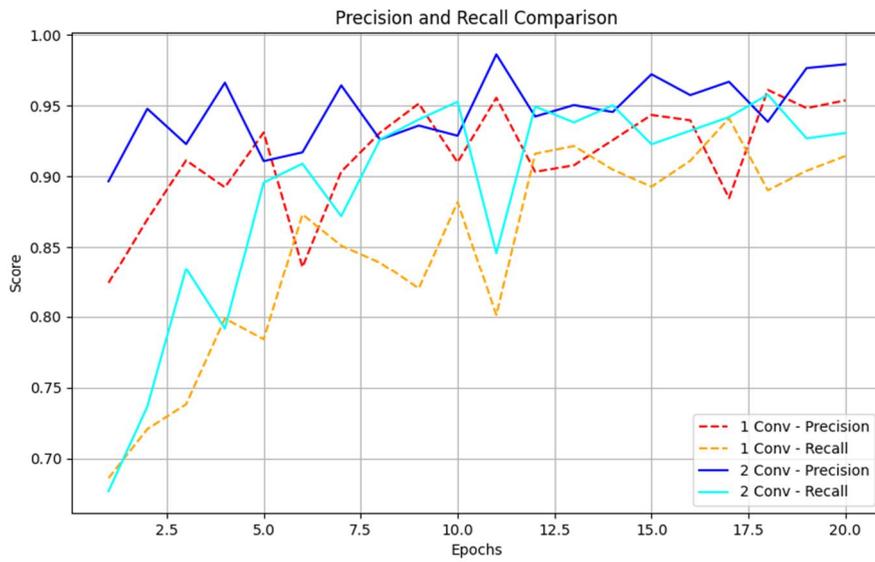


図 4.5 : 精度と再現率変化図

4.3 ネットワークのバリエーション

ネットワークのバリエーションにおいて、まずモデルにとって最適な畳み込みカーネルの数を見つける必要がある。セクション 5.1 の実験結果（2層の畳み込み深さを持つ TCN ブロック構造）に基づき、カーネル数を変更する実験を行いました。前述のように、モデル構築段階では、畳み込みカーネルの数が 16 のときに性能が最適であると仮定した。この仮説を検証するために、カーネル数をそれぞれ 4、8、16、32 に調整し、比較実験を実施した。

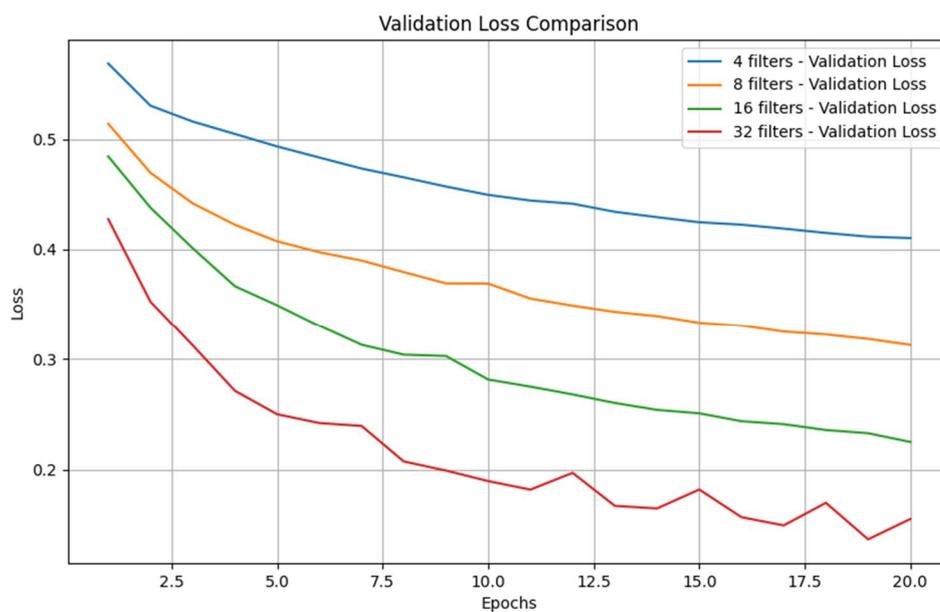


図 4.6：損失関数変化図 2

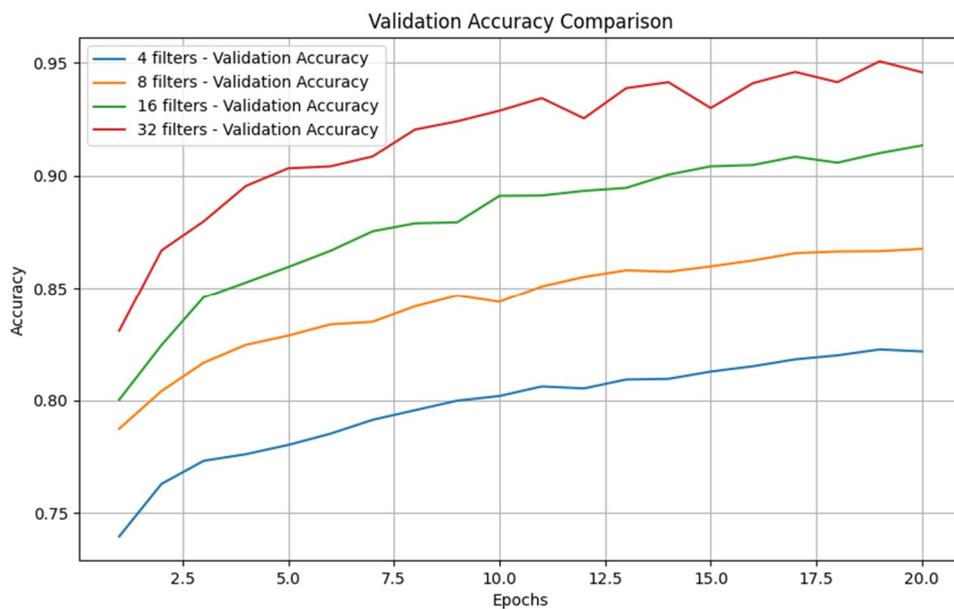


図 4.7 : 正確率変化図 2

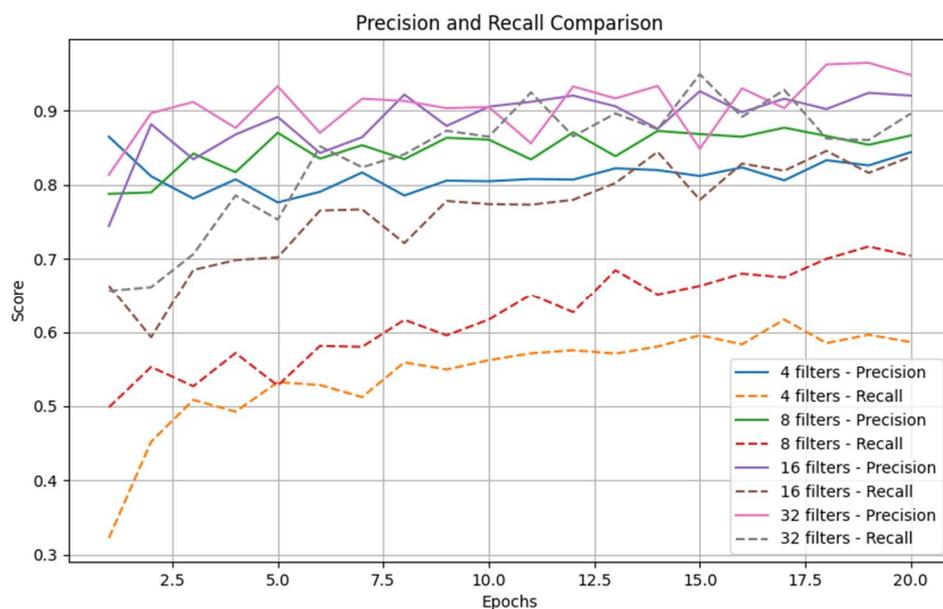


図 4.8 : 精度と再現率変化図 2

結果として、カーネル数が 16 および 32 のときに損失関数がより速く収束することが分かった。また、カーネル数 32 の場合、モデルの精度はトレーニング

プロセス中に着実に向上し、最終的な結果は他のカーネル数よりも優れた 98% となった (図 4.11 参照)。したがって、本研究では、より良い性能を得るためにカーネル数を 32 に設定したモデルを使用した。

第5章 おわりに

5.1 結果のまとめと従来方法との比較

モデルの性能を評価するために混同行列 (Confusion Matrix) を作成した。混同行列は、分類モデルの予測結果と実際の値を比較することで、モデルの分類精度を視覚的かつ定量的に評価するために使用される。データ処理した時、階層サンプリングを用いて、データを 70% のトレーニングセットと 30% の検証セットに分割し、`stratify=labels` を設定することで、分割後のトレーニングセットおよび検証セットのラベル分布が元のデータと一致するようにした。本研究では、モデルの性能を検証するため、テストデータを用いて予測を実施した。各サンプルに対して転倒 (1) または非転倒 (0) を分類し、その予測結果を実際のラベルと比較した。この比較結果を基に混同行列を作成した。

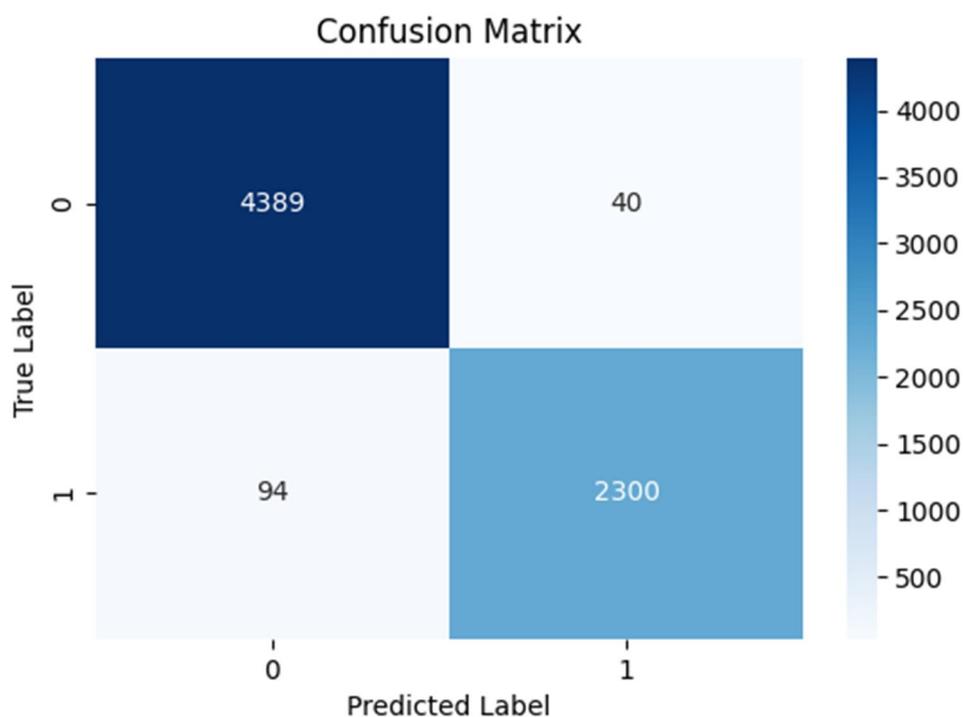


図 5.1 : MDATCN 混同行列

混同行列の各セルは以下の通りです：

True Negative (4389 件)：非転倒を正しく非転倒と予測した件数。

False Positive (40 件) : 非転倒を誤って転倒と予測した件数。

False Negative (94 件) : 転倒を誤って非転倒と予測した件数。

True Positive (2300 件) : 転倒を正しく転倒と予測した件数

Accuracy の計算方法は 4.2 のように、

$$\{Accuracy\} = \frac{TP + TN}{Total\ Samples} = \frac{2300 + 4389}{6823} \approx 0.9803$$

まとめに言うと、SVM、多モーダル CNN、単モーダル CNN、LSTM、TCN ベースラインネットワーク及び CNN+LSTM ハイブリッドモデルと比べて、MDATCN は自分のデータセットで最もよい性能をもらった。具体的に言うと正確率 0.9803 を到達した。

モデル	accuracy
SVM	0.705
CNN(加速度)	0.876
CNN	0.943
TCN	0.965
CNN+LSTM	0.959
MDATCN	0.9803

表 5.1: 関連研究正確率

そして、私別のベースラインモデルの計算時間、結果は表 5.2 のように、LSTM のモデルに比べて、MDATCN の速度はもっとはやくになる。でも、やっぱり CNN と普通 TCN より、やや遅くなる。まず CNN に比べて、MDATCN はもっと多くのコンボリューション層がある、従って計算量もおおくなる。そして普通 TCN より、MDATCN は余得な注意力に関する層があるので、計算時間が少し増加した。

モデル	MDATCN	Tcn	Cnn	Lstm	Cnn+lstm
times	0.032	0.03	0.028	0.036	0.039

表 5.2: 関連研究計算速度

5.2 重さの計算

マルチヘッドにより、加速度角速度角度三つのモダリティの重みの分布を計算できる。図 5.1 により、縦軸はアテンション重みの範囲（0 から 1）を表し、横軸はそれぞれのモダリティを分類する。

加速度計モダリティのアテンション重みの分布は、ほぼ 0 から 1 の全範囲にわたっており、中央値は 0.6 付近に位置している。これは、モデルが加速度計の情報を主要な判断材料として活用していることを示唆している。

ジャイロスコープモダリティは、アテンション重みの分布が比較的狭く、約 0 から 0.6 の範囲に収まっている。その中央値は加速度計よりも低く、モデルにおける寄与度が中程度であることを示している。

角度モダリティは、アテンション重みが最も低い分布を示しており、値は下限付近に集中している。また、外れ値（ウィスカー上の離散点）が確認され、このモダリティの中央値は 3 つの中で最も低く、モデルにおける影響力が限定的であることがわかるけど、特定の場所でモデルに補充の作用がある。

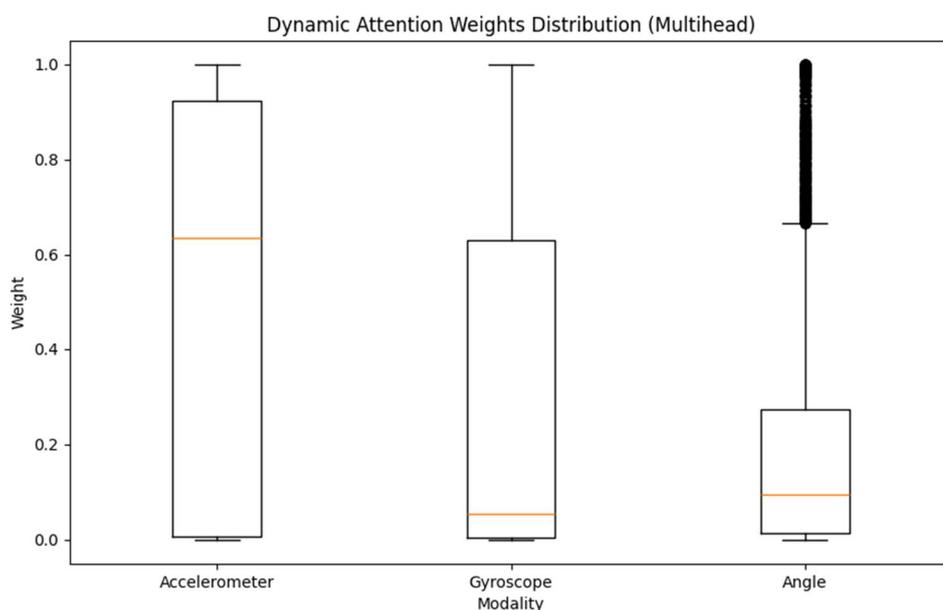


図 5.2 : 注意力重さ分布図

5.2 結論

転倒検出は人間の行動認識分野の重要な課題である。この問題に対して、従来の畳み込みニューラルネットワーク (CNN) と再帰ニューラルネットワーク

(RNN) の限界を克服するために、動的多モーダル注意メカニズムと時間畳み込みネットワーク (TCN) に基づく革新的なモデルを提案した。5層の因果畳み込みを含む TCN 構造を構築し、マルチヘッド注意メカニズムを融合することにより、多モーダル信号間のタイミング依存関係と特徴相互作用のモデリング能力を大幅に向上させた。実験の結果、提案した動的多モード注意力 TCN モデルは転倒検出タスクにおいて優れた検出性能を示した。モデルは動的重み生成モジュールと特徴融合戦略により、異なるモーダル信号の重みを適応的に割り当てることができ、モデルのロバスト性と柔軟性を高めることができる。単一モーダル方法 (例えば加速度信号のみを使用) と比較して、多モードモデルは検出精度と耐ノイズ能力の面で顕著な優位性を示している。さらに、LSTM などの再帰ニューラルネットワークとの比較実験は、本モデルの安定性、計算効率、および長シーケンスデータ処理能力における優位性を検証した。

5.3 将来の仕事

将来の研究では、次の点から最適化あるいは拡張を継続できる。まず、実際のデータセットの作成である。一般的に致命的な転倒は老人の間で発生するので、高齢者に関するデータが非常に必要である。しかし、現実生活における転倒行為の希少性とデータ取得の複雑性 (データプライバシー、意外性などの問題) のため、将来の研究では、シミュレーションデータと実際のデータとの間により強い相関を構築し、実際のシーンにおけるモデルの適用性を向上させることを模索すべきであると思う。そして、多モーダルデータの融合を強化し、現在のデータは単純な類似モーダル間の融合にすぎず、将来的には圧力センサー、温湿度センサーなどのより多くのセンサー入環境センサー、または心拍数、酸素飽和度などのバイオセンサに融合することを試みることができる、その他にも、モデルの軽量化と配置を改善し、現在のモデル再組み込み機器上の配置も非常に困難な剪定 (pruning)、量子化 (quantization) などの技術を用いてモデルパラメータを削減する。エッジ計算の最適化などの方法を用いて、スマートフォンやウェアラブルデバイスなどのエッジデバイスに計算を分布させ、クラウド計算への依存を減らすこともできる。

参考文献

- [1] World Health Organization. *Falls*. World Health Organization, <https://www.who.int/news-room/fact-sheets/detail/falls>. Accessed 8 Jan. 2025.
- [2] N. Pannurat, S. Thiemjarus and E. Nantajeewarawat, "Automatic fall monitoring: A review", *Sensors*, vol. 14, no. 7, pp. 12900-12936, 2014.
- [3] Y. Zigel, D. Litvak and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls", *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2858-2867, Dec. 2009.
- [4] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment", *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274-1286, Nov. 2012.
- [5] N. Pannurat, S. Thiemjarus and E. Nantajeewarawat, "Automatic fall monitoring: A review", *Sensors*, vol. 14, no. 7, pp. 12900-12936, 2014
- [6] Rougier C., Meunier J., St-Arnaud A., and Rousseau J. (2011b). Robust video surveillance for fall detection based on human shape deformation. *IEEE Trans. Circ. Syst. Video Technol.* 21, 611–622. doi: 10.1109/TCSVT.2011.2129370.
- [7] Djelouat H., Baali H., Amira A., and Bensaali F. (2017). "CS-based fall detection for connected health applications," in 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME) (Beirut: IEEE), 1–4. doi: 10.1109/ICABME.2017.8167540
- [8] Rougier C., Auvinet E., Rousseau J., Mignotte M., and Meunier J. (2011a). "Fall detection from depth map video sequences," in International Conference on Smart Homes and Health Telematics (Montreal: Springer), 121–128. doi: 10.1007/978-3-642-21535-3_16
- [9] Wang, X., Ellul, J., & Azzopardi, G. (2020). Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, 7, 71.
- [10] Hori T., Nishida Y., Aizawa H., Murakami S., and Mizoguchi H. (2004). "Sensor network for supporting elderly care home," in *Sensors*, 2004,

Proceedings of IEEE (Vienna: IEEE), 575–578. doi: 10.1109/ICSENS.2004.1426230

- [11] Mann S: *Wearable Computing*. http://www.interaction-design.org/encyclopedia/wearable_computing.html
- [12] El-Bendary N., Tan Q., Pivot F. C., and Lam A. (2013). Fall detection and prevention for the elderly: a review of trends and challenges. *Int. J. Smart Sens. Intell. Syst.* 6. doi: 10.21307/ijssis-2017-588
- [13] Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, 35, 68-80.
- [14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, et al., "Attention Is All You Need", *NIPS*, 2017.
- [15] Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [16] Yhdego, H., Li, J., Paolini, C., & Audette, M. (2021, December). Wearable sensor gait analysis of fall detection using attention network. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 3137-3141). IEEE.
- [17] Li, Q., Stankovic, J. A., Hanson, M. A., Barth, A. T., Lach, J., & Zhou, G. (2009, June). Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In 2009 sixth international workshop on wearable and implantable body sensor networks (pp. 138-143). IEEE.
- [18] Y. Yan and Y. Ou, "Accurate fall detection by nine-axis IMU sensor," 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), Macau, Macao, 2017, pp. 854-859, doi: 10.1109/ROBIO.2017.8324524.
- [19] Zhang, T., Wang, J., Xu, L., & Liu, P. (2006). Fall detection by wearable sensor and one-class SVM algorithm. In *Intelligent Computing in Signal Processing and Pattern Recognition: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006* (pp. 858-863). Springer Berlin Heidelberg.
- [20] Makino, K., Lee, S., Bae, S., Chiba, I., Harada, K., Katayama, O., ... & Shimada, H. (2021). Simplified decision-tree algorithm to predict falls for community-dwelling older adults. *Journal of clinical medicine*,

10(21), 5184.

- [21] Liu, C. L., Lee, C. H., & Lin, P. M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert systems with applications*, 37(10), 7174-7181.
- [22] A. H. Fakhruddin, X. Fei and H. Li, "Convolutional neural networks (CNN) based human fall detection on Body Sensor Networks (BSN) sensor data," 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 2017, pp. 1461-1465, doi: 10.1109/ICSAI.2017.8248516.
- [23] Zheng Y., Liu Q., Chen E., Ge Y., and Zhao J. L., Exploiting multi-channels deep convolutional neural networks for multivariate time series classification, *Frontiers of Computer Science*. (2016) 10, no. 1, 96–112.
- [24] Lee S.-M., Yoon S. M., and Cho H., Human activity recognition from accelerometer data using convolutional neural network, 2017 iee international conference on big data and smart computing (bigcomp), 2017, Jeju, South Korea, 131–134..
- [25] Torti, E., Fontanella, A., Musci, M., Blago, N., Pau, D., Leporati, F., & Piastra, M. (2019). Embedding recurrent neural networks in wearable systems for real-time fall detection. *Microprocessors and Microsystems*, 71, 102895..
- [26] Zebin T., Sperrin M., Peek N., and Casson A. J., Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks, 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2018, Honolulu, HI, USA, 4385–4388.
- [27] Zheng L., Li S., Zhu C., and Gao Y., Application of IndRNN for human activity recognition: the Sussex-Huawei locomotion-transportation challenge, *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, London, England, 869–872.
- [28] Lin, C.-B., Dong, Z., Kuan, W.-K., & Huang, Y.-F. (2021). A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Applied Sciences*, 11(1), 329. <https://doi.org/10.3390/app11010329>.

- [29] Yang, G., Zhao, J., & Guo, J. (2021, March). Research on Fall Detection Algorithm Based on CNN and LSTM. In 2021 3rd International Conference on Natural Language Processing (ICNLP) (pp. 190-195). IEEE.
- [30] Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1), 6-23.
- [31] Du, J., Li, W., Lu, K., & Xiao, B. (2016). An overview of multi-modal medical image fusion. *Neurocomputing*, 215, 3-20.
- [32] Prakash, A., Chitta, K., & Geiger, A. (2021). Multi-modal fusion transformer for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7077-7087).
- [33] Chudasama, V., Kar, P., Gudmalwar, A., Shah, N., Wasnik, P., & Onoe, N. (2022). M2fnet: Multi-modal fusion network for emotion recognition in conversation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4652-4661).
- [34] Abro, I. A., & Jalal, A. (2024, November). Multi-Modal Sensors Fusion for Fall Detection and Action Recognition in Indoor Environment. In *2024 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETEECTE)* (pp. 1-6). IEEE.
- [35] Bai S., Zico Kolter J., and Koltun V., Trellis networks for sequence modeling, 2018,
- [36] WIT Motion. (n.d.). Product documentation. Retrieved January 9, 2025, from <https://wit-motion.yuque.com/wumwnr/docs/syg9sx>
- [37] Khalifa, S., Lan, G., Hassan, M., Seneviratne, A., & Das, S. K. (2017). Harke: Human activity recognition from kinetic energy harvesting data in wearable devices. *IEEE Transactions on Mobile Computing*, 17(6), 1353-1368.
- [38] Maarit Kangas, Antti Kantilla, Per Lindgren, Ilkka Winblad and Timo Jämsä, "Comparison of low-complexity fall detection algorithms for body attached accelerometers", *Gait & posture*, vol. 28, no. 2, pp. 285-291, 2008.

