

機械学習を用いた時系列マルチモーダル起床動作予測 IoT システムの高精度化

代表研究者 森 本 尚 之 三重大学 大学院工学研究科 准教授

1 研究の背景

近年、少子高齢化や核家族化の進行に伴い、在宅での見守りや健康管理の必要性が高まっている。とりわけ高齢者や独居者においては、日常生活リズム、特に起床時刻および就寝時刻を把握することが、身体状態や生活状況の変化を検知するための重要な指標となる。高齢者は、起床直後にベッドからの転落や転倒などの事故が発生するリスクが高く、これが骨折や長期入院につながるおそれがある。そのため、起床状態を事後的に検知するだけでなく、事前に予測することで、適切な見守りや安全確保の介入を可能にする技術が強く求められている。このような技術は、医療・介護の現場だけでなく一般家庭への応用も期待されており、実用化が強く望まれている。

睡眠状態や起床状態を推定するためのさまざまな手法がこれまでに提案されており、たとえばリストバンド型加速度センサを用いた睡眠検知 [1], [2]、圧力センサを用いた動作検知 [3]、RGB カメラや赤外線カメラを用いた非接触モニタリング [4], [5] などがある。これらの各手法にはそれぞれ利点がある一方、実用化にあたっては、プライバシー保護と設置・運用の容易さを特に考慮する必要がある。

この観点から、Ogawa ら [6] は、赤外線センサを用いて低解像度の温度画像を取得し、それを深層学習モデルに入力することで、非接触かつプライバシーに配慮した形で起床状態を予測するシステムを提案している。この手法は、高解像度の映像データに依存せずにセンシングを行えるため、プライバシーを保持できるという利点を有する。しかし、被験者全体を含む赤外線画像をそのまま入力として用いる場合、毛布の有無や厚さ、被験者位置のずれ、周囲温度の変動といった要因がノイズとして作用し、予測精度の低下やモデル学習の不安定化を招く可能性がある。ここから、どの空間領域を入力として選択するかが、予測精度と計算効率の双方にどのような影響を及ぼすのかという問題が生じる。このトレードオフを明らかにすることは、実運用に向けた重要な一歩である。

したがって本研究では、Ogawa らが提案した既存の起床予測フレームワークに関心領域 (Region of Interest: ROI) 抽出を導入することで、予測精度と計算効率の双方の向上を目指す。特に、本システムを IoT アプリケーションとして実装する場合、利用可能なデバイスは一般に小型であり、計算資源およびメモリ資源に制約がある。そのため、推論の高速化とモデルの軽量化は極めて重要である。ROI 抽出によって入力画像サイズを縮小することで、起床検知に必要な本質的情報を保持したまま、計算コストを大幅に低減できる可能性がある。言い換えれば、精度を損なうことなく実行時間を短縮できれば、実用的な IoT デバイス上でのリアルタイム動作が可能となる。さらに、赤外線画像を複数の領域に分割して個別に分析することで、被験者のどの身体部位 (例えば、頭部、体幹部、脚部) が起床予測に強く関与し、どの部位の影響が小さいのかを定量的に評価することが可能となる。これは、モデルの解釈性を高めるだけでなく、今後のセンサ配置やデータ活用設計に有用な知見を与える。本研究で想定するシステムの全体図を Fig. 1 に示す。

本研究では、深層学習モデルの学習過程および計算効率に着目する。具体的には、赤外線画像に ROI 抽出を適用した場合の予測性能および実行時間への影響を検討する。さらに、実環境運用では対象者が単独で存在するという前提が崩れる可能性があるため、周辺条件の把握を目的として、Wi-Fi CSI を用いた複数人滞在状態の検出についても補完的に検討した。その結果、起床予測のような比較的細かな動作の識別には限界がみられた一方で、人の滞在状況のような、より粗い粒度の推定には活用の可能性があることが示唆された。本報告書では、これらの検討のうち、主として赤外線画像に対する ROI 抽出の導入とその効果に関する結果について述べる。本研究の貢献は、以下のとおりである。

- ROI 抽出が、マルチモーダル起床予測における予測性能と計算効率の双方に及ぼす影響を定量的に明らかにした。
- 中間領域の ROI と決定レベル融合モデルを組み合わせることで、毛布条件の違いに対しても最も安定した性能が得られることを確認した。
- ROI 抽出により、精度を維持したまま推論時間を大幅に短縮できることを示し、資源制約のある IoT 環

境への適用可能性を確認した。

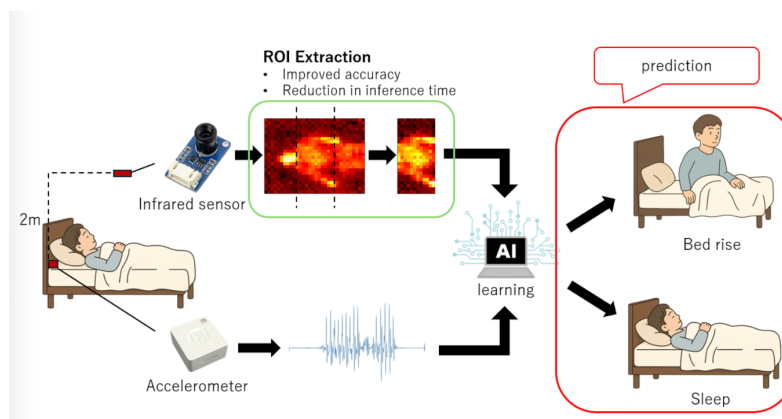


Fig. 1. 本研究で想定するシステムの全体図

2 関連研究

睡眠状態推定に関する研究では、加速度センサや赤外線センサを用いた非接触センシング技術が広く検討されてきた。Morawski ら [4] は、低解像度赤外線センサ (16×16 画素) を用いて高齢者の行動を分類し、非接触かつ低コストなセンシングの有効性を示した。同様に、Chen ら [5] は赤外線による温度分布に基づいて睡眠状態を推定し、ウェアラブル機器を必要としない睡眠モニタリングの実現可能性を報告している。Tang ら [7] は、赤外線画像を用いた睡眠姿勢の CNN ベース分類手法を提案し、深層学習に基づく非接触手法の可能性を示している。

これらの研究の流れを踏まえ、Ogawa ら [6] は、赤外線センサと加速度センサを組み合わせたマルチモーダル起床予測システムを提案した。このモデルでは、赤外線画像 (24×32 画素) と加速度データを入力とし、CNN によって空間的特徴を抽出し、LSTM によって時系列的特徴を処理している。さらに、毛布なし・薄い毛布・厚い毛布の 3 条件で学習を行うことで、異なる寝具環境に適応した起床分類を実現している。しかしながら、このモデルは毛布や背景領域を含む画像全体を処理対象としているため、学習に不要な情報が影響を及ぼす可能性があり、計算効率や推論時間の面で課題が残されている。

こうした課題に対処するため、近年では、精度を維持しつつ効率を向上させる手法として、ROI 抽出が注目されている。ROI 抽出は、入力画像中の有益な領域のみを選択することで、予測性能を損なうことなく計算コストを削減する手法である。医用画像分野では、Khan ら [8] が、診断精度を維持したまま計算効率を改善する ROI 限定型の深層学習手法を提案している。また、Sun ら [9] は、重要領域を自動的に強調する注意機構ベースの ROI 抽出ネットワークを導入し、高精度なセグメンテーションを実現した。さらに、軽量 CNN に基づく検出器の中には、二段階構成の枠組みで ROI 機構を利用するものもあり、モバイル端末や組み込みデバイス上においても高精度な物体検出を達成し、効率的なエッジコンピューティングに貢献している [10]。

ROI 抽出の応用は、睡眠関連研究にも広がっている。Yang ら [11] は、fMRI データを ROI 単位に分割することで、睡眠段階における脳活動遷移を詳細にモデル化した。Chen ら [12] は、咽頭 MRI データから ROI ベースのラジオミクス特徴を抽出し、閉塞性睡眠時無呼吸症候群の重症度分類性能を向上させた。Jakkaew と Onoye [13] は、赤外線カメラを用いて温度変化の大きい領域を自動抽出し、非接触呼吸検知を実現している。Kamon ら [14] は、小児の睡眠モニタリングにおいて、ベッド領域のみを ROI として抽出することで、動作検知の安定性を向上させた。Siddiqi ら [15] は、新生児睡眠モニタリングにおける脳波チャネルの寄与を分析し、少数のチャネルに絞っても精度を維持しつつ計測系を簡素化できることを示した。この手法は、画像ベースセンシングにおける ROI 解析と概念的に対応する、空間的指針に基づく特徴抽出プロセスとみなすことができる。

これらの研究は、ROI 抽出が不要な背景情報を効果的に除去し、有意なセンサデータに学習を集中させることで、計算効率と予測精度の両立に有効であることを示している。特に、低解像度赤外線画像に対しては、

ROI ベース処理によって計算負荷を軽減し、リアルタイム推論を可能にできるため、IoT デバイスへの応用に適していると考えられる。そこで本研究では、Ogawa らのマルチモーダル起床予測フレームワークを基盤として ROI 抽出を導入し、毛布条件の違いに対しても安定して動作する、軽量かつ頑健な起床予測モデルの実現を目指す。

3 研究方法

本研究では、赤外線センサと加速度センサを用いるマルチモーダル起床予測モデルに関心領域 (ROI) 抽出を導入し、その効果を定量的に評価した。本節では、ROI 抽出およびマルチモーダル融合手法に関する比較実験を含め、モデル構成、データセット、使用ツール、実行環境、および実験条件について述べる。

3-1 モデル構成

提案モデルでは、赤外線画像および加速度データから CNN を用いて特徴を抽出し、それらを統合した上で、LSTM 層により時系列的変化を学習する。LSTM は内部メモリ状態を保持することで短期的な時間依存性を効果的に捉えることができ、本研究の起床予測課題には十分適している。赤外線フレームからの空間特徴抽出には、畳み込み層、ReLU 活性化関数、および最大プーリング層から構成される標準的な CNN を用いた [16]。

本研究では、Transformer ベースのモデルのようなより複雑なアーキテクチャではなく、LSTM を採用した。その理由は、LSTM の方がパラメータ数および計算コストの両面で大幅に軽量であるためである [17]。これは、低遅延かつ限られたハードウェア資源のもとで動作することが求められる、IoT 指向の本システムに適している。また、提案モデルで CNN と LSTM を組み合わせた主たる理由は、画像データをそのまま LSTM に入力すると、必要なメモリセル数が画素数に比例して増加してしまうためである。これによりモデルパラメータが大幅に増加し、推論時間が長くなり、覚醒直前の起床タイミングを予測するという本研究の目的に対して不利となる可能性がある。この問題に対処するため、提案モデルではまず各センサモダリティに対して CNN を適用し、フレーム単位の特徴を抽出することで入力次元を削減している。

加速度データには一次元 CNN を、赤外線画像には二次元 CNN を適用した。これらの CNN の出力を結合し、LSTM 層へ入力することで、二値分類 (睡眠 / 起床) を行った。出力層には、確率的予測のためにシグモイド活性化関数を用いた。モデル学習には PyTorch 2.0 を使用し、損失関数には二値交差エントロピー、最適化手法には Adam を採用した。

3-2 データセットと実験条件

本研究で使用したデータセットは、Ogawa ら [6] により収集されたものであり、赤外線データおよび加速度データから構成される。被験者は 20 代男性 7 名であり、赤外線センサ (解像度: 24×32 画素) および 3 軸加速度センサを用いて、睡眠行動および起床行動を記録した。毛布条件として、「no (毛布なし)」「thin (薄い毛布)」「thick (厚い毛布)」の 3 条件を設定し、各条件について同一の手順でデータを収集した。赤外線画像の例を Fig. 2 に示す。

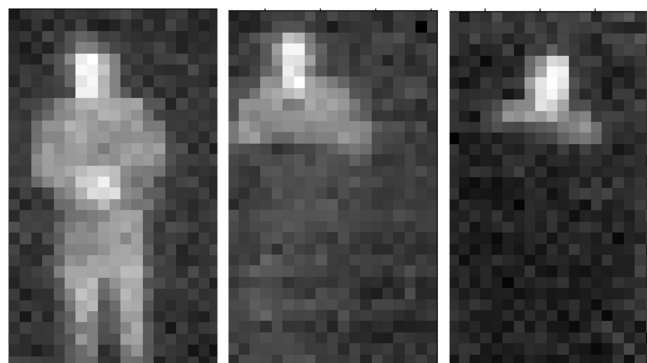


Fig. 2. 赤外線画像データの例 (左: 毛布なし、中央: 薄い毛布、右: 厚い毛布)

各試行は、5 秒間の初期姿勢フェーズ、10 秒間の姿勢変化フェーズ、5 秒間の起床フェーズから構成され、合計 20 秒であった。最後の 5 秒間の起床フェーズでは、被験者は実際に起き上がるか、あるいは睡眠状態を維持する。続くこの 5 秒間の起床フェーズにおいて、実際に起き上がったか、あるいは睡眠を継続したかに基づき、その直前の 10 秒間の姿勢変化フェーズに対してラベルを付与した。

3-3 ROI 抽出手法

本研究では、毛布条件の違いによって赤外線画像内に現れる空間的差異に着目し、どの領域が起床検知に最も寄与するかを検討した。赤外線画像を縦方向に上部・中部・下部の 3 領域に分割して、それぞれに対応する ROI を定義し、各領域ごとのデータセットを作成した。さらに、ROI 抽出を行わない全体画像（24×32 画素）をベースラインとして用いた。したがって、比較実験のために 4 種類の入力形式を用意した。ROI 抽出前後の赤外線画像の例を Fig. 3 に示す。

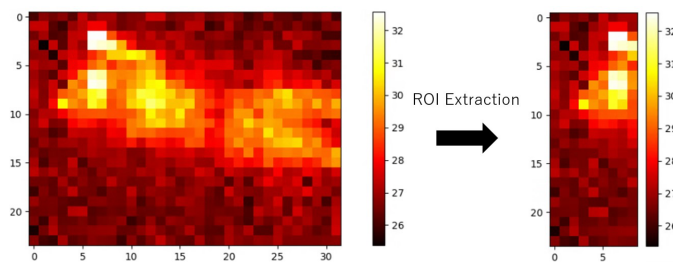


Fig. 3. ROI 抽出前後の赤外線画像の比較（左：24×32 の全体画像、右：24×9 の切り出し ROI）

このように領域分割を行った理由は以下のとおりである。上部領域は通常、毛布に覆われにくく、主として頭部や顔周辺の温度変化を捉える。下部領域は被験者の脚部に対応し、毛布のずれなどの動きによる温度変化を反映しやすい。一方、中部領域は一般に毛布で覆われる体幹部に相当し、体温変化が比較的安定かつ連続的に表れやすい。各領域は異なる種類の温度情報および動作情報を捉えると考えられるため、本研究では、それぞれの領域が起床予測精度にどのように寄与するかを比較した。

3-4 融合手法と評価指標

各 ROI に対して、3 種類のマルチモーダル融合手法 [18] を比較した。

- ・Feature Fusion: 赤外線データおよび加速度データから CNN によって抽出された特徴ベクトルを結合し、LSTM へ入力する手法である。これは早期融合あるいは表現レベル融合に相当し、モデルが両モダリティの統合特徴空間を学習する。

- ・Data Fusion: 赤外線データと加速度データを単一モデルに同時入力し、共同学習を行う手法である。これは中間融合に位置づけられ、共有されたモデルパラメータを通じて、生データからモダリティ間の関係性を直接学習する。

- ・Decision Fusion: 赤外線データ用モデルと加速度データ用モデルをそれぞれ独立に学習し、その出力確率を後段で統合する手法である。これは後期融合に相当し、最終的な予測スコアのみを統合するため、計算コストが低くなる傾向がある。

これらの融合手法は、すべての ROI 条件 (full, upper, middle, lower) およびすべての毛布条件 (no, thin, thick) のもとで検証し、合計 36 通りの実験条件を設定した。モデル性能の評価には、Accuracy、F1-score、学習時間、および推論時間を用いた。

3-5 実験環境

実験環境としては、Python 3.11、PyTorch 2.0、および NVIDIA GeForce RTX 3060 を用いた。実験結果の信頼性を高めるため、すべての条件について 100 回の学習を実施し、その平均値を最終評価値とした。再現性を担保するために、100 個の固定乱数シードを順次適用し、データセット分割およびモデル初期化を制御

した。この方法により、乱数初期化やデータ分割の偏りに起因する性能変動を抑制し、ROI 条件および融合手法間の安定した比較を可能とした。

4 結果と考察

本研究では、赤外線画像に対する ROI 抽出が、起床予測モデルの性能および計算効率の双方にどのような影響を及ぼすかを評価した。画像サイズ (full, upper, middle, lower) と毛布条件 (no, thin, thick) の各組合せについて、モデル学習を 100 回実施し、Accuracy、F1-score、ならびに学習時間および推論時間の平均値を比較した。

評価対象としたマルチモーダル統合手法は、Feature Fusion、Data Fusion、Decision Fusion の 3 種類である。Table I に Accuracy および F1-score の比較結果を、Table II に学習時間および推論時間の比較結果を示す。各行における最良値は太字で示している。以下では、各 ROI (full, upper, middle, lower) について、性能と効率の両面から結果を分析する。

TABLE I. ROI および毛布条件ごとの Accuracy と F1-score (100 回実行の平均)

ROI	Condition	Feature Fusion		Data Fusion		Decision Fusion	
		Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Full (24×32)	no	0.637	0.629	0.639	0.662	0.638	0.614
	thin	0.646	0.643	0.642	0.639	0.673	0.683
	thick	0.744	0.735	0.710	0.723	0.715	0.698
Upper ROI (24×9)	no	0.658	0.633	0.635	0.599	0.655	0.631
	thin	0.641	0.664	0.656	0.668	0.657	0.667
	thick	0.687	0.690	0.672	0.677	0.666	0.670
Middle ROI (24×12)	no	0.641	0.638	0.651	0.641	0.664	0.637
	thin	0.604	0.585	0.607	0.594	0.657	0.670
	thick	0.717	0.714	0.683	0.690	0.735	0.735
Lower ROI (24×11)	no	0.637	0.624	0.636	0.639	0.643	0.640
	thin	0.644	0.673	0.638	0.635	0.669	0.678
	thick	0.773	0.768	0.731	0.733	0.772	0.774

4-1 Full Image (24 × 32)

全体画像を用いた場合、thick 条件では Feature Fusion が最も高い Accuracy および F1-score を示し、thin 条件では Decision Fusion が最良の性能を示した。一方、no 条件では Data Fusion がわずかに優れていたが、3 手法間の差は全体として小さかった。学習時間および推論時間に関しては、Decision Fusion が他の手法より大幅に高速であり、すべての条件において最小値を記録した。

4-2 Upper ROI (24×9)

上部領域では、no 条件において Feature Fusion が最良の精度を示し、thin 条件では Decision Fusion、thick 条件では再び Feature Fusion が最良の性能を示した。特に no 条件では、Feature Fusion が 0.658 という最高 Accuracy を示した。学習時間および推論時間の観点でも、Feature Fusion が最も短時間で処理を完了した。これは、上部領域の実験において、Feature Fusion のネットワーク構成が比較的軽量であり、より複雑なモデルに比べて必要な計算資源が少なかったためであると考えられる。

TABLE II. ROI および毛布条件ごとの学習時間と推論時間 (100 回実行の平均)

ROI	Condition	Feature Fusion		Data Fusion		Decision Fusion	
		Training [s]	Inference [s]	Training [s]	Inference [s]	Training [s]	Inference [s]
Full (24×32)	no	35.1	0.055	84.5	0.110	25.7	0.047
	thin	36.5	0.056	84.0	0.111	27.6	0.047
	thick	42.2	0.058	102	0.112	30.6	0.047
Upper ROI (24×9)	no	17.1	0.035	70.9	0.098	22.1	0.039
	thin	20.4	0.034	88.3	0.094	23.9	0.040
	thick	21.8	0.034	90.2	0.095	23.6	0.040
Middle ROI (24×12)	no	23.5	0.037	93.3	0.109	16.7	0.030
	thin	22.7	0.037	83.4	0.099	16.6	0.030
	thick	24.2	0.038	95.1	0.098	18.1	0.030
Lower ROI (24×11)	no	20.3	0.034	85.8	0.109	23.0	0.040
	thin	20.5	0.034	77.9	0.101	24.0	0.038
	thick	24.5	0.033	95.4	0.099	27.2	0.040

4-3 Middle ROI (24×12)

中部領域は、全体として最も安定した傾向を示した。毛布条件 no、thin、thick のすべてにおいて、Decision Fusion は最も高い Accuracy を示し、F1-score についても thin 条件および thick 条件で最良、no 条件でも他手法と同程度の性能を示した。また、Decision Fusion は学習時間・推論時間の両面でも他手法より短く、特に thin 条件では 16.6 秒という最短学習時間を示した。これらの結果は、中部領域が動作情報および温度変化情報を豊富に含んでおり、Decision Fusion が空間特徴抽出と時系列モデリングを効果的に統合できたことを示唆している。この領域は、呼吸や体幹動作に伴う微細な温度変化が表れやすいため、加速度データとの統合も性能向上に寄与したと考えられる。

4-4 Lower ROI (24×11)

下部領域では、thick 条件において Feature Fusion と Decision Fusion の双方が比較的高い Accuracy (0.773、0.772) および F1-score (0.768、0.774) を示した。しかし、no 条件および thin 条件では、thick 条件に比べて精度が低く、この領域で利用可能な情報が毛布の被覆状態の影響をより強く受けることが示唆された。これらの結果から、脚部の動きは、特に厚い毛布条件下において有効な識別手がかりとなることが示された。学習時間および推論時間は上部領域・中部領域よりやや長かったものの、Feature Fusion が最短の処理時間を示し、Decision Fusion も高い精度を維持しつつ高い効率性を示した。

4-5 総合的な考察

同一 ROI 内で融合手法を比較すると、上部領域では Feature Fusion が優位であった一方、中部領域および下部領域では Decision Fusion が優位となる傾向がみられた。中部領域と Decision Fusion の組合せは、すべての毛布条件において、上部領域と Feature Fusion の組合せより高い Accuracy を示した。さらに、中部領域と下部領域における Decision Fusion を比較すると、thick 条件では下部領域の方がわずかに高性能であり、thin 条件でもわずかに上回った。しかし、下部領域は毛布条件による性能変動が比較的大きかったのに対し、中部領域は毛布条件に依存せず安定した精度を維持した。加えて、中部領域と Decision Fusion の組合せは、下部領域における Feature Fusion または Decision Fusion よりも短い学習時間・推論時間を実現した。

また、元の全体画像 ROI と比較した場合でも、中部領域と Decision Fusion の組合せは同等の精度を維持しつつ、必要な推論時間を全体画像モデルの約 60～65%にまで削減した。以上より、中部領域 ROI と Decision Fusion の組合せが、精度と処理時間のバランスに最も優れ、毛布条件の違いに対しても安定した性能を実現することが明らかとなった。

4-6 Decision Fusion に関する統計解析

以上の結果を踏まえ、本研究では Decision Fusion の結果に着目し、full ROI と middle ROI に限定して統計解析を実施した。

Table III には、Decision Fusion における full ROI 設定と middle ROI 設定の統計的比較結果を示す。各毛布条件について、Accuracy および推論時間の平均値、標準偏差、95%信頼区間を算出した。また、Table III には、Decision Fusion モデルにおける full ROI と middle ROI の対応のある統計比較結果として、Wilcoxon 検定の p 値および効果量指標 r (順位二系列相関係数) も併せて示している。

毛布なし条件および厚い毛布条件では、middle ROI は full ROI よりも統計的に有意に高い Accuracy を示した ($p < 0.01$, $r \approx 0.3$)。一方、薄い毛布条件では、middle ROI が full ROI を上回るという仮説は棄却され ($p = 0.98$, $r = -0.175$)、むしろ full ROI の方が良好な性能を示す傾向が確認された。推論時間に関しては、middle ROI はすべての毛布条件において full ROI よりも有意に高速であった ($p < 0.01$, $r \geq 0.96$)。これらの結果は、middle ROI が毛布条件によっては full ROI と同等またはそれ以上の精度を実現しつつ、推論時間を一貫して短縮できることを示している。

次に、100 回の実行から得られた混同行列の各要素の平均値を算出し、平均混同行列ならびに平均 FNR および平均 FPR を求めた (起床 = positive、睡眠 = negative)。Table IV に示すように、3 種類の毛布条件 (No / Thin / Thick) のすべてにおいて、middle ROI は一貫して低い FPR を示した一方、FNR は full ROI と同程度、あるいはやや高い傾向を示した。

TABLE III. Full ROI と Middle ROI の統計的比較 (Decision Fusion)

Condition	Metric	ROI	Mean	SD	95% CI	<i>p</i> -value	<i>r</i> (effect size)
No Blanket	Accuracy	full	0.638	0.066	[0.625, 0.651]	<0.01	0.313
		middle	0.664	0.075	[0.649, 0.679]		
	Inference Time	full	0.0467	0.0016	[0.0464, 0.0470]	<0.01	0.98
		middle	0.0302	0.0053	[0.0291, 0.0312]		
Thin Blanket	Accuracy	full	0.673	0.056	[0.662, 0.684]	0.98	-0.175
		middle	0.657	0.055	[0.646, 0.668]		
	Inference Time	full	0.0465	0.0022	[0.0461, 0.0470]	<0.01	0.96
		middle	0.0303	0.0043	[0.0295, 0.0312]		
Thick Blanket	Accuracy	full	0.715	0.079	[0.699, 0.730]	<0.01	0.293
		middle	0.735	0.069	[0.722, 0.749]		
	Inference Time	full	0.0467	0.0027	[0.0462, 0.0472]	<0.01	0.98
		middle	0.0300	0.0043	[0.0292, 0.0309]		

5 結論と今後の課題

本研究では、赤外線データおよび加速度データを用いた起床予測において、ROI 抽出およびマルチモーダル統合手法について検討した。赤外線画像は、上部・中部・下部・全体の4領域に分割し、3種類の統合手法 (Feature Fusion、Data Fusion、Decision Fusion) を比較することで、毛布条件の違いに応じた有効性を評価した。

実験の結果、ROI 抽出を導入することで、予測精度を同程度に維持しながら、学習および推論の双方における入力画像次元と計算時間を大幅に削減できることが確認された。特に、middle ROI と Decision Fusion の組合せは、精度、処理時間、ならびに毛布条件の違いに対する安定性の観点から、全体として最もバランスの取れた性能を示した。また、lower ROI は厚い毛布条件下で高い精度を示しており、寝具を介した温度変化や脚部動作が有効な識別手がかりとなる可能性が示唆された。これらの結果は、適切な ROI を選択することが、推論速度と予測精度の両立に有効であることを示している。

一方で、本研究にはいくつかの課題が残されている。第一に、データセットは20代男性7名のみを対象として収集されたものであり、年齢、性別、身体的特性の多様性が十分ではない。そのため、介護現場における主たる対象であり、より緩慢またはばらつきの大きい起き上がり動作を示す可能性のある高齢者に対して、本手法がどの程度一般化可能であるかは未解明である。

第二に、本研究では被験者依存評価を採用している。今後は、未知被験者を対象とした検証を含む被験者非依存の評価設定についても検討し、提案手法の汎化性能を評価する必要がある。

TABLE IV. Full ROI および Middle ROI における平均混同行列と誤り率 (Decision Fusion) (なお、FN、FP、TN、TP は 100 回の実行で得られた各混同行列要素の平均値である)

Condition	ROI	FN	FP	TN	TP	FNR	FPR
No Blanket	full	20.41	20.17	35.53	35.89	0.3713	0.3681
	middle	21.67	15.97	39.73	34.63	0.3905	0.2905
Thin Blanket	full	16.03	20.61	34.42	40.94	0.2835	0.3772
	middle	18.50	19.90	35.13	38.47	0.3298	0.3670
Thick Blanket	full	17.84	14.10	41.46	38.60	0.3180	0.2569
	middle	15.81	13.83	41.73	40.63	0.2836	0.2527

第三に、本手法では固定的な ROI 設定を前提としているため、被験者の就寝姿勢やベッド上の位置が変化した場合には、ROI のずれが生じる可能性がある。このような、あらかじめ定義した ROI と実際に有用な領域との不一致は、実環境において予測精度の低下を招くおそれがある。この課題に対しては、被験者位置に応じて温度分布や特徴マップをもとに ROI を自動調整する Dynamic ROI Selection の導入を今後の課題とする。この手法により、必要な領域のみに注目することで、予測精度と計算効率のさらなる向上が期待される。

第四に、本研究では、代表的な失敗事例に対する定性的なエラー分析を行っていない。こうした定性的検討は、モデルの挙動理解を補完する重要なアプローチであると認識しており、今後取り組むべき重要課題と位置づけている。

最後に、本研究で用いたモデルは CNN-LSTM アーキテクチャのみに限定されている。今後、Transformer のような別の時系列モデルや、従来型機械学習手法についても評価を進めることが、さらなる性能向上に向けた重要な検討課題である。

以上の改善に加え、赤外線データや Wi-Fi CSI を含む新たなセンシング技術の活用可能性について、検討および学会発表を引き続き進めることで、実環境での継続運用に向けて、より高精度かつ汎用性の高い起床予測システムの実現を目指す。

【参考文献】

- [1] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.
- [2] Y. Nam, Y. Kim, and J. Lee, "Sleep monitoring based on a tri-axial accelerometer and a pressure sensor," *Sensors*, vol. 16, no. 5, p. 750, 2016.
- [3] R. Adachi, Y. Yamauchi, N. Shimoi, and K. Nakasho, "Optimal positioning of sensors in a bed monitoring system using RFID," in *Proc. IEEE International Conference on Consumer Electronics–Taiwan*, pp. 429–430, 2024.
- [4] I. Morawski, W.-N. Lie, and J.-C. Chiang, "Action prediction using extremely low-resolution thermopile sensor array for elderly monitoring," in *Proc. IEEE International Conference on Image Processing*, pp. 984–988, 2021.
- [5] Z. Chen and Y. Wang, "Sleep monitoring using an infrared thermal array sensor," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2019*, vol. 10970, pp. 409–418, SPIE, 2019.

- [6] K. Ogawa and N. Morimoto, “Bed rise prediction IoT system under varying blanket conditions using multimodal machine learning,” in *Proc. IEEE International Conference on Consumer Electronics-Asia*, pp. 78–81, IEEE, 2024.
- [7] K. Tang, A. Kumar, M. Nadeem, and I. Maaz, “CNN-based smart sleep posture recognition system,” *IoT*, vol. 2, no. 1, pp. 119–139, 2021.
- [8] I. J. Khan, M. F. B. Amin, M. D. S. Deepu, H. K. Hira, A. Mahmud, A. M. Chowdhury, S. Islam, M. S. H. Mukta, S. Shatabda, A. D. N. Initiative, *et al.*, “Enhanced ROI guided deep learning model for alzheimer’s detection using 3D MRI images,” *Informatics in Medicine Unlocked*, p. 101650, 2025.
- [9] L. Sun, W. Shao, D. Zhang, and M. Liu, “Anatomical attention guided deep networks for ROI segmentation of brain MR images,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2000–2012, 2019.
- [10] P. Mittal, “A comprehensive survey of deep learning-based lightweight object detection models for edge devices,” *Artificial Intelligence Review*, vol. 57, no. 9, p. 242, 2024.
- [11] F. N. Yang, D. Picchioni, J. A. de Zwart, Y. Wang, P. van Gelderen, and J. H. Duyn, “Reproducible, data-driven characterization of sleep based on brain dynamics and transitions from whole-night fMRI,” *eLife*, vol. 13, p. RP98739, 2024.
- [12] Y. Chen, H. Xiao, M. Huang, Y. Zheng, X. Dong, and G. Chen, “Predictive modeling of obstructive sleep apnea using pharyngeal MRI radiomics and clinical data,” *Journal of Clinical Sleep Medicine*, pp. jcsn–11706, 2025.
- [13] P. Jakkaew and T. Onoye, “Non-contact respiration monitoring and body movements detection for sleep using thermal imaging,” *Sensors*, vol. 20, no. 21, p. 6307, 2020.
- [14] M. Kamon, S. Okada, M. Furuta, and K. Yoshida, “Development of a non-contact sleep monitoring system for children,” *Frontiers in Digital Health*, vol. 4, p. 877234, 2022.
- [15] H. A. Siddiq, M. F. Qureshi, A. Khurshid, Y. Xu, L. Wang, S. F. Abbasi, C. Chen, and W. Chen, “EEG electrode setup optimization using feature extraction techniques for neonatal sleep state classification,” *Frontiers in Computational Neuroscience*, vol. 19, p. 1506869, 2025.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.
- [17] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [18] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Impact of ROI Extraction from Infrared Images on Prediction Performance and Efficiency in Multimodal Bed Rise Prediction	5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)	2025 年 12 月
マルチモーダル起床予測における赤外線画像 ROI 抽出領域と予測性能の関係	情報処理学会 第 88 回全国大会	2026 年 3 月
Wi-Fi CSI を用いた個室内の複数人滞在検出による異常検知	情報処理学会 第 88 回全国大会	2026 年 3 月