「マルチメディア，分散，協調とモバイル
(DICOMO2019) シンポジウム」 令和元年7月

行動認識における表現学習モデルと個人依存に関する考察

長谷川 達人[1,a)]　越野 亮[2]

概要：センシングによる行動認識研究が広く行われているが，個人依存性が問題になることがある．センサデータを用いた行動認識について幅広くサーベイした結果，CNN（Convolutional Neural Network）を用いた表現学習モデルによる行動認識について十分な検討がなされていなかった．そこで本研究では，画像認識分野で研究が進んでいる CNN モデルをベースに，行動認識における表現学習モデルの有効性の検証実験を行った．行動認識のベンチマークデータセットに対して，HC（Hand-crafted）特徴量を用いた DNN，シンプルな CNN モデル，AlexNet，FCN，VGG，ResNet，SENet 等 10 種類のモデルに対して，訓練データの多様性を変化させて 6 種類，ランダム性を考慮して 10 セットで，計 600 回深層学習モデルを訓練し推定精度検証を行った．その結果，訓練データに被験者を多く確保できる場合には，SE-VGG が最も高い精度を達成することを明らかにした．更に，訓練データを十分に確保できない場合には HC 特徴量が有効に働くことや，HC 特徴量は個人依存の影響を比較的強く受けることも明らかにした．

Consideration of Representation Learning Models and Individual Dependency in Activity Recognition

TATSUHITO HASEGAWA[1,a)]　MAKOTO KOSHINO[2]

1. はじめに

人間の動作や行動をセンシングにより認識する研究分野がある．近年のウェアラブルデバイスや IoT 機器の普及により，身近にセンサ機器がありインターネットに接続されていることが当然となった．そのため，多彩なセンサデータを蓄積することが容易になり，ビッグデータとして日々増え続けている．一方，センサデータは一般的に波形データであることが多く，計測結果を人間が直接観察しても，それを解釈することは容易ではない．そこで，センサデータに自動で解釈を加える技術の研究が盛んに行われている．

身につけた加速度センサの計測結果から，行動を示す特徴量を抽出し，機械学習等の手法によって利用者がどのような行動（歩行や走行，着座等）を行っていたのかを自動で分類する行動認識は，特に盛んに研究が行われている．行動認識研究には，動画像からの人物検出と行動推定を行う研究もあるが，本稿では加速度等のセンサデータを用いた所持者の行動認識研究に焦点を当てて議論を行う．最近では，利用者が日々身につけているスマートフォンやスマートウォッチを用いた研究が盛んである．また，運転中の動作や，家庭内の動作，ヘルスケア等の様々なドメインに特化した行動認識研究もある．

センシングによる行動認識研究では個人依存が問題になることが多い．例えば，20 歳の健常男性と高齢者では多くの場合に歩き方や動作の傾向が異なるため，汎用的な行動認識モデルを開発する場合はトレードオフを取らざるを得ない．汎用的なモデルで特定個人の行動認識を行う場合よりも，特定個人に特化したモデルで行動認識を行う場合の方が，認識精度は高いことが確認されている [1]．では，特定個人に特化したモデルを開発すれば良いかと言えば，実運用を考えると容易ではない．個人に特化したモデルを開発するためには，特定個人に関するアノテーションされたデータセットを用意する必要があるためである．実運用を考えると，新規利用者は必ず最初にアノテーションされたデータセットを準備せねばならないことになる．これはユーザビリティの低下につながり，現実的な運用ではない．

[1] 福井大学大学院工学研究科
Graduate School of Engineering, University of Fukui
[2] 石川工業高等専門学校電子情報工学科
Department of Electronics and Information Engineering, National Institute of Technology, Ishikawa College
[a)] t-hase@u-fukui.ac.jp

表 1　センサを用いた行動認識に関する関連研究
Table 1　Related works on activity recognition using sensors.

Ref.	Target labels	Device	Sensor	Position
[1]	Sit, Lie, Stand, Walk, Jog, UpStairs. Watch TV, Stretch, Fold laundry, Brush teeth, Carry items, Scrub, Ride elevator, Work on PC, Eat or drink, Read, Bicycle, Training, Vacuuming, Ride escalator	Original	Acc.	Right wrist, Right ankle, Right hip, Left arm, Left thigh
[2]	Sit, Lie, Stand, Walk, Others	Original	Acc.	Chest, Both thighs
[3]	Sit, Lie, Stand, Dynamic, Others	Original	Acc.	Chest, Rear thigh
[4]	Stand, Walk, Jog, StairsUp, StairsDown, Vcuum, Brush teeth, Situps	Original	Acc.	Pelvic region
[5]	Sit, Lie, Stand, Walk, Jog, Bicycle, Ascend, Kneel, Descend, Guestures (Chop, Throw, Punch, Clockwize, Anticlockwise, Jump, Kick)	Original	Acc.	Right wrist, Waist, Right ankle
[6]	Walk (include StairesUp and down), Watch TV, Brush teeth, Shave, Hairdryer, Vacuum, Toilet flush & hand wash, Dish-wash, Iron	Original (Phone like)	Acc. & Microphone	Chest pocket
[7]	Sit, Stand, Walk, Jog, StairsUp, StairsDown	Smartphone	Acc.	Trouser pocket
[8]	Sit, Stand, Walk, Jog, StairsUp, StairsDown	Smartphonel	Acc. & Gyro. & Grav.	Trouser pocket
[9]	Sit, Stand, Walk, Jog, StairsUp, StairsDown Bicycle, Type, Write, Coffe, Eat, Talk, Smoke	Smartphone, Smartwatch	Acc. & LinearAcc. & Gyro.	Pocket, Wrist
[10]	Guesture (finger, hand or arm)	Smartwatch	Acc. & Gyro.	Wrist

多くの研究では汎用的なモデルの推定精度を向上させる手法を模索している．個人依存による推定精度低下を低減する仕組みについては，行動認識に限らず様々なコンテキストアウェアネス研究において重要課題である．

一方，特に画像認識分野で表現学習 [11] に関する研究が進んでいる．従来は人間の経験に基づいて，生データから特徴量を設計していたが，与えられたデータセットに基づいて特徴表現自体を学習する手法が表現学習である．行動認識分野においても徐々に表現学習が取り入れられており [12]，深層学習を用いた表現学習手法が比較検討されている．しかし，画像認識分野ほど徹底的な議論はなされておらず，特に個人依存に関して未解明な部分も多い．

本研究では，行動認識のベンチマークデータセット HASC[13] を用いて，様々な CNN モデル構造を適用して行動認識を行い，モデル構造に対する推定精度を考察する．特に，統一条件によって，訓練データに含まれる被験者の多様性を制限し，個人依存による推定精度変化に対する考察を深める点が本稿の特色である．本研究により，今後の深層学習を用いた行動認識研究におけるベースラインモデルを明らかにすることを目的とする．

本研究の主な貢献は以下の通りである．

- 行動認識研究について幅広くサーベイし，行動認識で用いられている認識手法を体系的に取りまとめる．特に，従来の特徴抽出から機械学習を行う手法だけでなく，近年の深層学習による認識手法まで取り扱う．
- CNN モデルを複数種類実装し行動認識問題に適用する．ベンチマークデータセットを用いて，行動認識に適した表現学習モデル構造を明らかにする．訓練データが充実しているケースから，していないケースを再現し，データセットの規模による影響も明らかにする．
- 個人依存による推定精度低下に関して考察する．関連研究では，個人依存を考慮しないケースも見受けられることから，本稿では一定の条件に統制した環境で，モデルごとの個人依存の影響の違いを明らかにする．

2. 行動認識研究

センサを用いた行動認識研究は，「認識対象」や「計測手法」，「推定手法」等のいくつかの方向性で分類することができる．「認識対象」はどの行動を認識するかであり，応用先に応じて対象が異なってくる．「計測手法」はどのような手法で推定のもとになる生データの計測が実現されるかであり，使用するセンサの種類やセンサの着用位置，着用方法等が異なってくる．「推定手法」は計測した生データからどのような手法で行動を推定するのかであり，前処理や推定アルゴリズムが異なってくる．本章では，これらの項目を幅広く調査した上で本研究の立ち位置を明確にする．

2.1 認識対象

行動認識研究において，特に応用分野や計測手法に着目して関連研究の調査を行った結果の一部を**表 1** に示す．表 1 より，多くの研究において認識対象となる行動は日常生活行動である．特に基本となる行動は，着座，横臥，直立，歩行，走行，階段上り，階段降りであり，今回の調査結果の大部分がこれを対象としているものであった．より

詳細な日常生活行動として，掃除機がけや歯磨き，食事，読書等を対象としている研究もある [1], [4], [6], [9]．これらの日常生活行動は一定時間の間，定常的に発生する動作である．一方で，所定のジェスチャを検出する研究も行われている [5], [10]．日常生活行動に対してジェスチャは開始時点から終了時点にかけて 1 セット実施されるような特定動作を意味する．そのため日常生活行動は一定の時間長のセンサデータをもとに行動を推定するのに対して，ジェスチャ認識は時間長が一定でないことや，発生タイミングがわからないといった特徴がある．

行動認識は医療や福祉に利活用することを目的とする研究が多い．文献 [2], [3] では，医用工学における動作分析を行う目的で，日常生活を常時自動記録できないかという課題に取り組んでいる．記録したセンサデータから，患者毎に行いがちな行動を分析したり，着座時間や移動動作にかけた時間を抽出しリハビリのアセスメントに応用できる．文献 [6], [8] では福祉への利活用を提言している．日常生活が推定できることで，独居高齢者の見守りや安否確認，生活行動の変化を検出することなどに応用できる．

他にも，行動認識を含むコンテキストアウェアネス研究は，検出した情報を様々なサービスに応用する事ができる．現実環境の事象をコンピュータが認識できることで，HCI（Human Computer Interaction）への応用による新しいインタフェースの開発や，LifeLog として記録し続けることで自身の行動把握や改善を促すサービスの開発等のコンシューマ向けサービスへの展開が期待できる．これだけにとどまらず，例えばスマートフォンが所持者の行動情報を記録し，それを集約することでマーケティングや渋滞情報の分析など，高次元な人間行動の分析が実現できる．

2.2 計測手法

表 1 より，大部分の行動認識研究では加速度センサを用いて行動のセンシングを行っている．表 1 の Original は加速度センサを用いた独自のデバイス（もしくは市販のシンプルな加速度センサ）を身体に着用させて計測を行っている研究である．特に，1990 年代から 2000 年代は現在ほど身の回りにコンピュータが普及していなかったことから試作的に実施されていた研究が多い．一方 2000 年代後半から，スマートフォンが急速に普及し始め，近年の行動認識研究ではスマートフォンで計測したデータを使用することが増えてきた．2010 年代ごろから，スマートウォッチを始めとするウェアラブルデバイスが徐々に普及しており，市販のデバイスが増加したことから，これを応用した研究が増えてきている [6], [7], [8], [9].

特に，スマートフォンは常日頃から利用者が身につけていると言う特性があり，更に電源や情報の記録，通信機能を独自で保有することから，行動認識研究と相性が良い．スマートフォンが普及したことにより，一般利用者にとっても日常生活行動の常時記録が実用的になった．スマートウォッチを含むウェアラブルデバイスも，スマートフォンほどではないものの広く普及している．こちらも常日頃から利用者が身に着けているという特性から，行動認識研究と相性が良い．スマートフォンはポケットやカバンに帯同されることが多いのに対し，ウェアラブルデバイスは着用箇所が限定されていると言う利点がある．すなわち，スマートフォンはどこに帯同されているかが利用者により変動するため，行動認識精度に悪影響を及ぼすことが知られている [14]．一方ウェアラブルデバイスは着用箇所が限定されているため，帯同場所によるゆらぎの影響が小さい．また，スマートウォッチであれば腕，スマートグラスであれば顔といったように，データ計測箇所がそれぞれの着用位置に特化されているという特性がある．したがって，スマートフォンは体全体の大まかな動き（歩行や走行等）を検出することはできるが，手指の細やかな動作や頭部動作などを検出することは難しい．一方で，ウェアラブルデバイスは着用箇所に特化した動作の検出に優れており，スマートウォッチであれば手の動作（タイピングや筆記，料理，食事等）の検出に優れ，スマートグラスであれば顔の動作（表情変化や精神状態）の検出に比較的優れている．

2.3 予測手法

2.3.1 機械学習を用いた行動認識

表 1 の大部分は機械学習を用いて行動認識を行っている．行動認識に関するサーベイ論文 [15], [16], [17] でも，網羅的に手法の概要が調査されている．本稿でも改めて，センサデータを用いた行動認識，コンテキストアウェアネス研究の多くが共通して採用する推定手法について説明する．

多くの研究では共通して**図 1** の手順でセンサデータから人間の行動を推定する手法を採用している．まず計測した生データ（Raw data）から，前処理（Pre-processing）を実施する．加速度センサだと微細なノイズが含まれやすいことから，ローパスフィルタを用いたスムージングが行われたり，通信や計測が不安定なことから発生する欠損値や異常値の補完処理が一般的に実施される．次に，前処理後のデータに対して時系列分割（Segmentation）を実施する．歩行や走行などの定常的な動作は，一定の間所定の波形が継続的に検出される．そこでセンサデータを一定の時間窓に分割し，その時間窓に対する行動を予測する問題に置き換える．この時間窓を，スライディングウィンドウ方式で時間方向にスライドさせることで，複数のインスタンスを生成する．時間窓のオーバーラップも許容する．その後，時間窓毎に特徴量抽出（Feature extraction）を実施する．特徴抽出しない場合サンプル数×センサ軸数が機械学習の入力となる．例えばフレームサイズ 256 で，3 軸加速度センサのみを用いた場合でも 768 次元となり，次元の呪い問題に陥る可能性が高い．そこで，人間の知見に基づいた

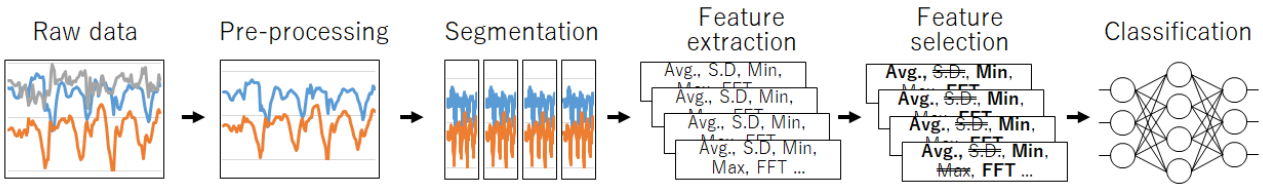

図 1　機械学習を用いた行動認識の一般的な処理手順

Fig. 1　General processing method of activity recognition using machine learning

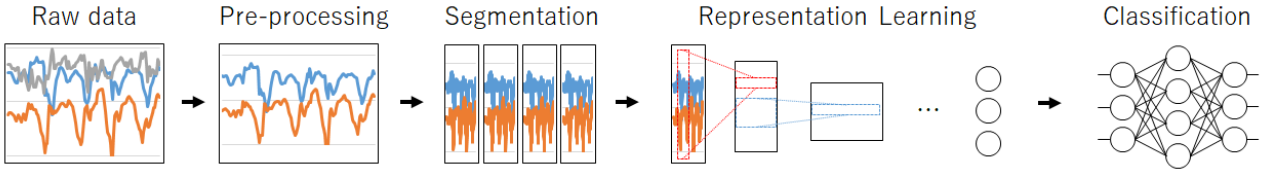

図 2　表現学習を用いた行動認識の一般的な処理手順

Fig. 2　General processing method of activity recognition using representation learning

HC（Hand-crafted）特徴量をセンサデータから抽出する処理を実施する．しかし，センサデータについては人間が目視で見ても特徴を判断しづらく，有益な特徴はあまり提案されていない．多くの研究では，各軸毎に平均や標準偏差等の基本統計量や，軸間の相関係数，尖度や歪度 [18]，ゼロ交差率（Zero-crossing rate）[19] を算出して特徴量とする．歩行や走行等の周期的な動作の検出時には FFT（Fast Fourier Transform）を実施後に，ピーク時の周波数やそのパワースペクトル値を特徴量とすることもある．特徴抽出後にも高次元のデータとなる場合，機械学習時に過学習により汎化性能が低下する可能性がある．その場合は，特徴選択（Feature selection）もしくは次元削減（Dimension reduction）を実施し，特徴量を厳選する事がある．特徴選択手法には，ANOVA の F 値に基づき上位 K 件の特徴を採用する SelectKBest の手法や，RandomForest[20] のようにモデルベースの特徴選択手法などがある．次元削減手法は，主成分分析（PCA : Principal Component Analysis）が有名である．最終的に，選択された特徴量と事前に手動で付与した正解ラベルの組み合わせをもとに，機械学習によって分類器を構築し，センサデータから行動の分類（Classification）を行う．分類器を構築する機械学習アルゴリズムも多くの手法が提案されており，決定木（C4.5）や，SVM（Support Vector Machine），Random Forest が有名である．なお，これらの手続きは研究により採用する手続きが異なったり，一部手順を省略したりすることがある．

2.3.2　深層学習を用いた行動認識

近年，深層学習に関する研究が盛んに行われている．深層学習はニューラルネットワークを深層化したもの全般を指す言葉であり，層の深いネットワークをうまく学習させることによって，高い表現力を獲得するモデルである．以前は，学習データセットが充実していない問題や，勾配消失により学習が進まない問題，計算コストの問題などにより上述した従来手法に遅れを取っていた．2006 年，RBMs（Restricted Boltzmann Machines）を多段に重ねた自己符号器（Auto-Encoder）が Hinton らによって提案され [21]，この時期を機に深層化したニューラルネットワークの研究が大いに発展した．画像認識の分野では，2012 年の ILSVRC（Imagenet Large Scale Visual Recognition Challenge）[*1] で優勝を果たした AlexNet[22] の登場がブレイクスルーとなった．前年には，従来から研究されていた画像特徴量の抽出から線形識別器による分類が行われていた．それに対して，AlexNet では 8 層の畳み込みニューラルネットワーク（CNN）を用いて，入力画像から特徴表現を直接学習する手法であり，2012 年の ILSVRC では 2 位以下に 10%以上の精度差をつけて大勝している．

行動認識研究に深層学習を導入するメリットは多くある．最も重要なメリットは特徴表現の自動抽出である．前節で述べたとおり，行動認識におけるこれまでの予測手法は，センサの観測データに対して前処理を施した後，特徴抽出を行い機械学習で予測する手法が大半であった．特徴抽出に着目すると，行動認識固有の知見に基づいて提案された特徴量は多くなく，基本統計量や波形データに関する特徴量を採用していることが多い．これは，多種多様な加速度波形が観測される中，行動を表現する固有の波形特徴を人間が発見することが容易ではないことに起因する．一方，深層学習を導入することで，**図 2** のように観測した生データから人間より優れたレベルの固有の特徴表現を自動抽出できる可能性がある．他のメリットとしては，モデルの柔軟性がある．深層学習で扱われるモデルは工夫次第で転移学習や半教師あり学習等の応用手法の利用が容易である．行動認識研究は，センサデータを獲得することが容易な一方，アノテーションが難しい事が多いため，モデルの柔軟性は大きなメリットとなりうる．

[*1] ILSVRC: http://www.image-net.org/challenges/LSVRC/

表 2　CNN を用いた行動認識研究
Table 2　Activity recognition studies using CNN.

Ref.	Architecture	Dataset	Evaluation
[23]	(Conv-Pool)*3	Original	Hold-out (cross subj.)
[24]	(Conv-Pool)*3+FC*2	Original, OPPOTUNITY [25], [26]	LOSO-CV
[27]	Conv+Pool+FC	Skoda[28], MHEALTH[29]	Hold-out (same subj.)
[30]	Conv+Pool+FC*2	OPPOTUNITY[25], [26]，Skoda[28]，Actitracker[31]	N-fold CV(same subj.)
[32]	(Conv-Pool)*3+FC*3	Original	10-fold CV(cross subj.)
[12]	(Conv-Pool)*3+FC	OPPOTUNITY[25], [26]，UniMiB SHAR[33]	LOSO-CV
[34]	(Conv-Pool)*2 + Conv+FC	OPPOTUNITY[25], [26]，Hand Gesture	Hold-out (same subj.)
[35]	(Conv-Pool)*3+FC	OPPOTUNITY[25], [26]，UniMiB SHAR[33], PAMAP2[36], [37]	Hold-out (same subj.)
[38]	Inception*4+GRU*2	OPPOTUNITY[25], [26]，PAMAP2[36], [37], UCI[39]	Hold-out (cross subj.)
[40]	Residual Bidir-LSTM	OPPOTUNITY[25], [26]，UCI[39]	Hold-out (cross subj.)
[41]	Inception+Hand-crafted+FC*2	UCI[39]	Hold-out (same subj.)
[42]	Dual Residual Network	OPPOTUNITY[25], [26]，UniMiB SHAR[33]	Unknown

デメリットとしては計算コストが高いという点が挙げられる．一般的に深層学習を用いた予測手法は従来の機械学習手法と比べ計算量が膨大であるため，スマートフォンやウェアラブルデバイス上でオンラインで動作するサービスに応用する際には，リアルタイム性とのトレードオフに関する議論が必要となる．他方で，末端のデバイスはセンシングと記録，通信機能のみを保持し，データの処理はクラウド上で一元管理する方法もある．例えばライフログやマーケティング活用のような，リアルタイムな推定が必要ないようなサービスの場合，クラウド上で一元管理する手法が採用できる．この場合，モデル構築に要する計算コストはさほど問題にならない．

深層学習を用いた行動認識に関する研究は，Wang らが幅広くサーベイしている [43]．シンプルな DNN（Deep Neural Network）から，RBM，SAE（Stacked Auto-Encoder），CNN，RNN（Rucurrent Neural Network）等の幅広いモデルを対象に研究事例の調査が行われている．行動認識に関するベンチマークデータセットについても調査されており，19 種のデータセットを紹介している．本研究は表現学習による個人依存への影響を考察することを目的としていることから，特に CNN モデルを採用している研究を Wang らのサーベイ [43] 以降に出版された論文も含めて調査した．(1) どのようなモデル構造を採用したか，(2) どのようなデータで評価したか，(3) 個人依存を適切に評価したかを基準に調査結果を**表 2** に集約した．

表 2 より，現在実施されている CNN を用いた行動認識研究では，Convolution と Pooling 層を 2〜3 層重ね，全結合層につなげるモデルを採用していることが多い [12], [23], [24], [27], [30], [32], [34], [35]．モデル構造は同じ中，データの取り扱い方が文献毎に少しずつ異なる点は興味深い．Ha らの研究 [27] では，複数の 3 軸加速度センサとジャイロスコープで観測された波形データを 2 次元データ（横：フレームサイズ，縦：センサ軸）として二次元畳み込みを行う手法を採用し，1 次元畳み込みの場合との比較を行っている．一方，Zeng らの研究 [30] では，センサの軸ごとに別々の Conv-Pool 層を学習させ，最終的に全結合する手法を採用している．Yang らの研究 [35] では，センサ毎（3 軸を 3 チャネルとする）に Conv-Pool 層を学習させ，最終的に全結合する手法を採用している．

我々が調査した範囲では，より高度なモデルを採用している研究 [38], [40], [41], [42] もいくつか見られた．Xu らの研究 [38] では，ILSVRC2014 で優勝した GoogLeNet[44] で採用している Inception module を多段に重ね，Gated Recurrent Unit[45] に接続する手法を提案している．類似する手法として，Zhao らの研究 [40] では双方向の LSTM[46] ブロックに対して，残差を学習する Residual 構造を導入した手法を提案している．その他，プレプリントのみの公開文献まで広げて調査すると，Dong らの研究 [41] では，Inception module で学習した特徴表現に加えて，HC 特徴量を結合する手法を提案している．Long らの研究 [42] では，広い範囲を見るネットワークと狭い範囲を見るネットワークを別々に学習し，最終的に全結合する手法を提案している．ILSVRC2015 で優勝した ResNet[47] で提案された Residual block を導入している点が特徴的である．

調査結果より，現在実施されている CNN ベースの行動認識研究にはいくつかの課題が見つかった．一つは画像認識分野ほど深い CNN モデル構造に関する議論がなされていないことである．前述の通り，少しずつ高度なモデルが検証されてきているものの，行動認識研究分野では現在進行系で研究が進んでいると言う印象を受ける．もう一つは個人依存に関する評価がされないことがあるという点である．これについては次節で議論する．

2.3.3　個人依存

センシングによる行動認識研究の特徴として，アノテーション済みデータが獲得しづらいことと，個人依存による推定精度の低下という問題がある．前者について，多くの

行動認識研究では，被験者の体の各所にセンサを取り付け，分類対象とする全ての行動を複数セット実施している．訓練データ作成時においては，観測したデータに加えて動画像を撮影するなどして，作業者の目視により波形データに対して正解ラベルを付与する．例えば数時間のセンサデータを複数人分アノテーションする作業を考えると，作業者の負担は膨大である．このように，データ収集作業自体の負担に加え，アノテーションにかかる負担があり，アノテーション済みデータの獲得が容易ではない．後者について，文献 [17] の今後の研究課題の Classifier flexibility として指摘されている通り，加速度データは人によって異なるため頑健な行動認識モデルの構築が求められる．例えば，子供と成人の歩行動作は大きく異なることから，観測される加速度波形も異なることが予測される．すなわち，事前に他者で訓練されたモデルで新規利用者の行動を推定する場合，新規利用者に特化した場合に比べて推定精度が低下することが多い．では，新規利用者に特化したモデルを訓練できればよいのだが，そのためには新規利用者のアノテーション済みのデータセットが必要となり，新規利用者に対する負担が大きい．そこで，他者で訓練されたモデルが利用者の違いに対して頑健であるほどよいモデルと言える．特にスマートフォンでデータを観測する場合，スマートフォンがどこに所持されるのかわからないこと（ズボンのポケットの中なのか，カバンの中なのか，手に持たれているのか等）や，利用者の服装が異なること（きつめのポケットなのか，余裕のあるポケットなのか等）に起因して，個人依存による推定精度低下が起こりやすい．

3. 行動認識のため CNN アーキテクチャ

行動認識における CNN を用いた表現学習モデルの有用性検証を行う上で，本稿で比較評価を行うモデル構造について述べる．本稿では HC 特徴量を用いた DNN（以降，DNN），関連研究 [12] で提案されているシンプルな CNN モデル（以降，Li2018），AlexNet，FCN（Fully Convolutional Network），VGG，ResNet の 6 種類である．それぞれのモデル構造を**図 3** に示す．これらのモデルはそれぞれのモデル名称をベースに実装しているが，1 次元センサデータを学習するためにモデル構造を一部変更している．

DNN では，従来の行動認識研究で一般的に用いられてきた HC 特徴量を抽出し，シンプルな 3 層の Dropout 付き DNN で行動認識を行うモデルである．深層学習手法と HC 特徴量を比較している文献 [12] に加えて，行動認識のサーベイ論文 [16], [17] を参考に，本研究では**表 3** に示す代表的な HC 特徴量を採用した．Li2018 では，入力データとして加速度センサの観測生データを与え，Convolution-Pooling を 3 セット繰り返した後に全結合し予測を行うモデルである．ここまでに説明した 2 つのモデル（DNN，Li2018）は従来研究で使われている代表的な手法であり，今後説明するモデルに対するベースラインとして取り扱う．Li2018 を含め，以降説明する全ての CNN モデルでは，入力を加速度センサの生観測データ（1 次元波形 256 サンプルを xyz 軸の 3 チャネル）とする．

本稿で比較検証を行う 4 つの CNN モデルについて説明する．AlexNet[22] は 2012 年の ILSVRC で 2 位以下に大差をつけて優勝を果たした，ディープラーニングの火付け役となった画像認識モデルである．当時有用性が示されつつあった活性化関数 ReLU[48] を導入したことや，Dropout による過学習の抑制等の工夫がなされている．VGG[49] は 2014 年の同コンテストで 2 位となった画像認識モデルである．これまで，比較的浅い層では大局的な情報を見るようにカーネルサイズやストライドを大きくしていたが，VGG では全層に渡って $(3, 3)$ の比較的小さなカーネルを用いる．図 3 左の ConvBlock に示すように，小さなフィルタを Pooling を挟まず連続で畳み込む事により，大きなフィルタと同等の効果を獲得する．VGG シリーズは層の深さに応じて異なるネーミングがなされているが，今回は図 3 に示す 16 層のネットワークを採用した．ただし，純粋な VGG をシンプルに 2 次元から 1 次元で実装しただけでは学習が収束しなかったことから，本稿ではフィルタサイズを小さくして VGG を模したネットワークを構築した．これらのモデルは現在もベースラインモデルとしてよく用いられていることから採用した．

FCN は画像のセグメンテーション分野の研究で提案されたモデル構造 [50] で，Pooling 層を用いずに畳み込み層と Global な Pooling 層を用いることで位置情報をより正確に保持したモデルである．位置情報を正確に扱うことから，時系列分類問題での有効性も議論されており [51], [52]，行動認識にも有効に働くと考え採用した．

ResNet[47] は 2015 年の ILSVRC で優勝した画像認識モデルであり，最大の特徴は Residual block を導入している点である．これ以降に提案されたモデルの多くが Residual 構造を採用していると言われているほど強力なモデルである．図 3 左の ResidualBlock に示すように，入力テンソルを直接伝えるパスと，畳み込みを連続で行うパスに分岐し最終的に足し合わせる．これにより入力 x に対して出力

表 3 本稿で採用した HC 特徴量の一覧（計 51 次元）
Table 3 List of Hand-crafted features we adopted in this paper (total 51 dimensions).

Domain	Features
Time	Mean, S.D., Variance, Min, First quartile, Median, Third quartile, Max, Interquartile Range (IQR), Mean Absolute Deviation (MAD), Root Mean Square (RMS), Mean Crossing Rate (MCR), Correlation between axes, Skewness, Kurtosis
Freq.	Spectral Energy, Spectral Entropy

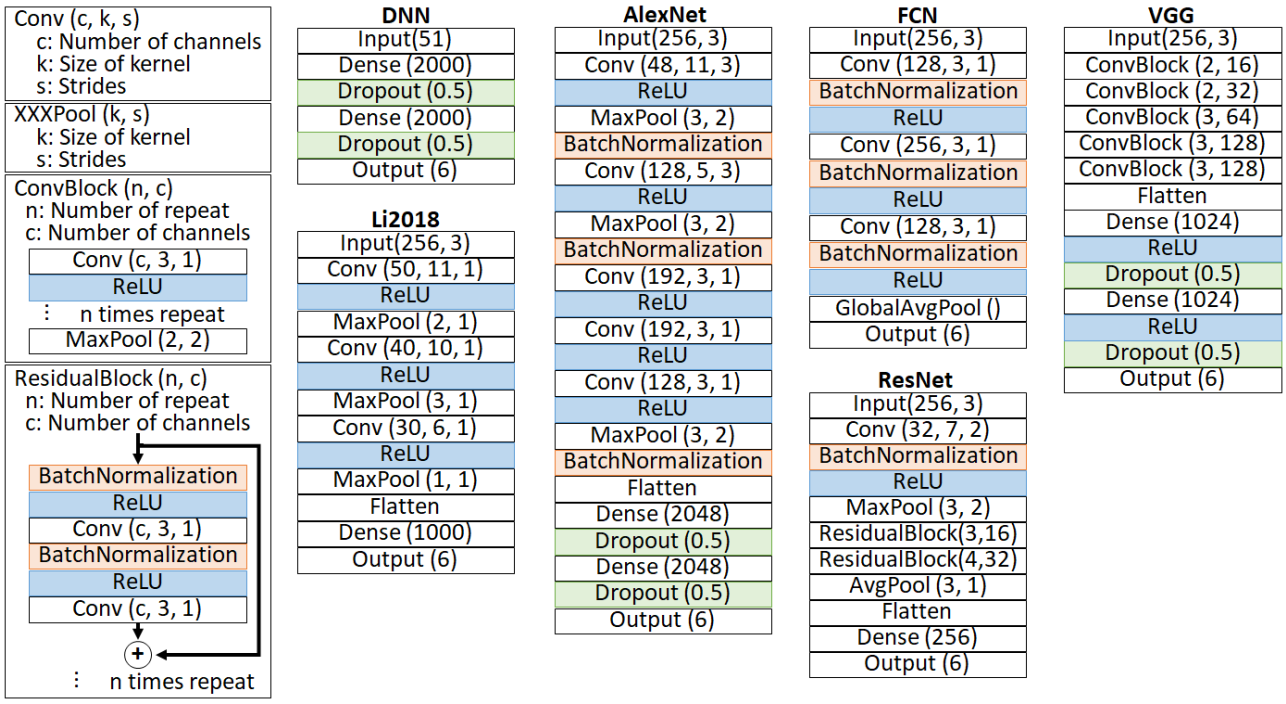

図 3　実験に使用する行動認識モデルのアーキテクチャ

Fig. 3　Activity recognition model architectures used in our experiments.

$z = F(x) + x$ を学習するため，入力に対する残差を学習する構造を実現している．ResidualBlock の導入により深層化が容易になったことで ResNet は高い表現能力を獲得している．今回は，層を深くしすぎること，フィルタが多すぎることで学習が収束しなかったことから，探索的に図 3 の 19 層の構造を採用した．

最後に，2017 年の ILSVRC で優勝した SENet (Squeeze-and-Excitation Networks) [53] で採用されている SEBlock について説明する．SENet は SEBlock の機構を様々なモデルに組み込んだネットワークを総称しており，Inception モジュールと組み合わせた SE-Inception モジュールや，Residual モジュールと組み合わせた SE-ResNet モジュールが文献内で紹介されている．SEBlock は一般的に畳み込み層の直後に挿入され，畳み込み後の各チャネルに適応的な重みを乗ずる役割を持つ．図 4 に示す通り，SEBlock のアーキテクチャは入力テンソルを直接伝えるパスと，強調するための適応重みを算出するパスに分岐し，最終的に入力に適応重みを乗ずる機構となっている．重み算出時にははじめに GlobalAveragePooling 層でチャネルごとに代表値を計算する．その後チャネル数をあえて減少させるボトルネックを作り，もとのチャネル数に戻すことによって重要な情報のみを強めるような適応重みを学習する．

本稿では上述した 4 つの比較モデルに対して，SEBlock 導入有無による推定精度変化の考察も執り行う．SEBlock ありの場合には，図 3 に示す構造の Conv 層の直後に SEBlock を配置することでこれを実現する．

4. 評価実験

表現学習モデルの有用性を評価するにあたり，行動認識のベンチマークデータセットを用いて，検証実験を行った．本章では，まずデータセットの概要を説明し，実施した実験の手順を説明する．その後，評価結果と考察を示す．

4.1 HASC データセット

HASC データセット [13] は，人間行動理解のための装着型センサによる大規模データベースの構築[*2]を目的とした非営利任意団体が提供する行動認識データセットである．基本行動 6 種類（停止，歩行，走行，スキップ，階段上り，

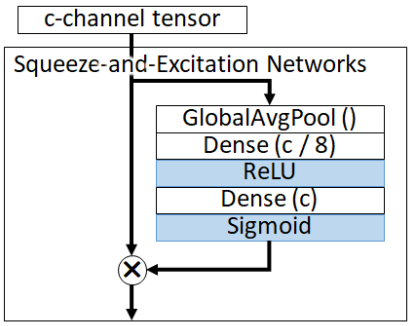

図 4　SENet (Squeeze-and-Excitation Network) で用いられる SEBlock のアーキテクチャ

Fig. 4　Architecture of SEBlock used in SENet (Squeeze-and-Excitation Network).

[*2] HASC: http://hasc.jp/

階段下り）のラベルがついた加速度，ジャイロ等のセンサデータがコーパスとして提供されている．iPhone3Gs 等のスマートフォンから，センサデバイス WAA-001 のような幅広い機器で計測されており，性別などの各種情報をメタ情報として保持している．

本稿では，2011 から 2013 年までのコーパスの BasicActivity より 297 名のデータから，加速度センサの生データのみを用いることとした．重力成分と加速度成分に分割することで推定精度が向上する可能性も指摘されているが，今回はすべての手法間で統一的に計測値の生データをそのまま使用する．前処理として，各計測ファイルから前後 5 秒を除去し，フレームサイズ 256 サンプル，ストライド 256 サンプルで時系列分割を行う．計測開始から端末の格納動作等の影響を取り除くため前後 5 秒でトリミングしている．計測機種や性別等のメタ情報は用いない．トリミング後に 1 フレーム以上データが取得できた 274 名全員のデータを例外なく採用した．

4.2 実験概要

モデルの評価は被験者を基準とした Hold-out 法により実施する．全データセットからランダムに 100 名を抽出し，検証用データセット D_{valid} とする．残ったデータからランダムに N 人を抽出し，訓練データセット D_{train} とする．すなわち，モデルの訓練時に評価用のユーザを含まず，他者のデータで学習したモデルで推定精度を評価する．本研究では，訓練用データの多様性に応じて個人依存による推定精度の影響がどの程度発生するのかを考察するため，N を可変とし，5 名，10 名，25 名，50 名，100 名とする．ランダム抽出によるばらつきを考慮し，実験は各 10 セット試行した上での算術平均で議論する．

検証モデルは前述の 6 種類であり，AlexNet, VGG, FCN, ResNet については SEBlock の有無も考察するため，合計 10 種類のモデルで検証を行った．最適化は Adam[54] を学習率 $= 0.001$，$\beta_1 = 0.9$，$\beta_2 = 0.999$ で採用し，損失関数はカテゴリカルクロスエントロピーとし，全て統一で 200 エポック学習した．D_{valid} に対する検証精度（Accuracy）を各エポックで算出しており，検証精度が最良だったものを評価指標とした．実運用では，検証精度が最大となるところで学習を停止することはできないが，今回最終エポックにおける精度と最良精度を比較した結果若干の差異しか見られなかった．最終エポックで議論する場合，偶然勾配が爆発したタイミングで 200 エポックを迎える事による精度低下をノイズ的に含んでしまうことから，本稿では最良精度を採用して議論を行う．

4.3 実験結果

D_{train} の人数を変化させたときの，各モデルの推定精度を図 **5** に示す．横軸が D_{train} に含まれる被験者数 N であり，縦軸が D_{valid} に対する検証精度である．全体の傾向に着目すると，当然ではあるが N が増加するとどのモデルも推定精度が向上する．訓練データが持つ被験者間の行動の多様性が増すことで，幅広い被験者の行動を予測できるようになり，個人依存による影響が低減できている．

予測モデル間の違いを考察するため N=100 に着目すると，モデル毎の推定精度は $VGG > Li2018 \geq AlexNet \geq DNN \geq ResNet > FCN$ という結果になった．いくつか興味深い点がある．一つは Li2018 が AlexNet の精度をわずかに上回るという点である．Li2018 に比べ AlexNet は若干深層なモデルだったが，N=100 での表現力は Li2018 がわずかに勝る結果となった．もう一つは，画像認識で現在主力な ResNet や，時系列分類で有効とされる FCN の精度が他に比べて低かったことである．特に ResNet は繰り返し数やフィルタサイズをいくつか探索的に検証しているが HC 特徴量を用いた DNN に及ばない結果となった．FCN は採用したモデルの層数が浅かったことが原因と推察するが関連研究 [51], [52] では有効な結果を示していただけに原因の追求が今後の課題である．SEBlock の有無に着目すると，必ずしも有効に働くわけではないが，FCN と VGG においては精度を向上させることが確認できる．これらの結果より，N=100 のように十分な訓練データセットを準備できる場合は，SE-VGG を採用することで 81.6%と最も高い精度を達成できることが明らかとなった．

N の変化による影響に着目すると，N=5 では僅差ではあるが DNN が最も高い精度を達成している．すなわち，訓練データセットが充実していない場合は表現学習を行うための十分なデータが確保できず，HC 特徴量の方が優れた特

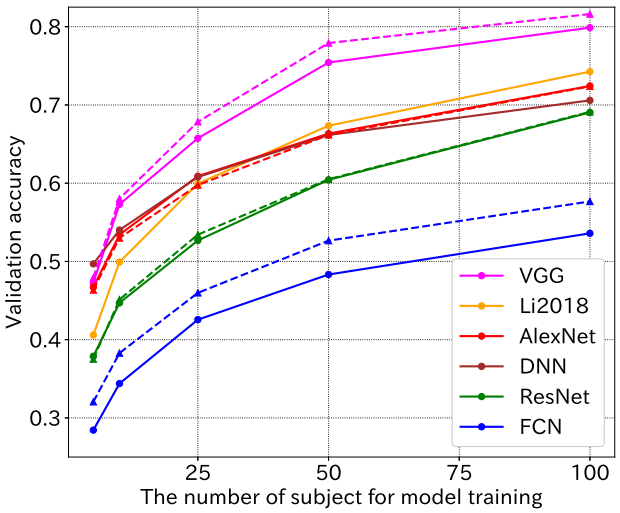

図 5 訓練に用いた被験者数に応じた推定精度の変化（実線：SEBlock なし，破線：SEBlock あり）

Fig. 5 Estimation accuracy versus the number of subjects for model training (solid line: without SEBlock, dashed line: with SEBlock).

徴表現となったと考えられる．一方で N が増加するにつれて CNN モデルの精度の方が上回る様子も確認できる．特に VGG は N=5 では DNN に劣っているものの，N=10 の段階で既に DNN を大きく上回っており，最終的に N=100 で 10%以上の精度向上を実現している．したがって，CNN を用いた表現学習モデルは訓練データの被験者数が極端に少ないケースでは HC 特徴量を用いた手法に劣る可能性があるが，ある程度の被験者を確保できる場合には大きく精度を改善することを明らかにした．

4.4 表現学習モデルと個人依存

個人依存について考察を行うため，自己データをある程度訓練データに含むことができるケースにおける推定精度を検証する．本来であれば D_{valid} の被験者各々について，自身のデータを用いてモデルを学習させ（もしくは既存モデルを Fine-tuning 等の手法で再学習させ），同じ被験者の別のデータで検証を行うことで，個人特化モデルの推定精度の検証が実施できる．しかし，D_{valid}100 名に対してこれを実施するには計算時間がかかりすぎることから，次の手順で個人特化モデルに類するモデルを検証した．

(1) D_{valid} に対して，被験者毎に時系列方向にデータを 2 分割する．ランダムスプリットは行わない．
(2) 2 分割したデータを用いて 2-fold CV（2-fold Cross-Validation）を行う．

この検証手法により，訓練時には D_{valid} 全てのユーザのデータから約半分のデータを訓練に用い，残り半分で検証を行う事ができる．そのため，訓練データに検証被験者自身のデータを含むことができ，自身のデータを訓練し個人依存の影響を低減させたモデルの検証が実現できる．ただし，完全な意味で検証被験者それぞれに個人特化したモデルでは無いため，さらなる精度向上の余地がある点には注意されたい．なお，本検証結果を以降 2-fold CV と呼び，前節で実施した結果を Hold-out と呼ぶ．

CNN モデルごとに個人依存度合いを比較した結果を**図 6** の箱ひげ図に示す．代表例としてベースラインモデル 2 種（DNN，Li2018）と，最も良い精度を達成した SE-VGG を並べた．緑の箱ひげ図が Hold-out 法による結果（自己データを訓練しない），黄色の箱ひげ図が 2-fold CV による結果（自己データを訓練する）である．縦軸は検証精度である．どの手法も個人依存による推定精度の低下が起こっていること，すなわち，自己データを学習し個人特化できることで推定精度を向上させられることが確認できる．

ここで，興味深い点が 2 点ある．一つは DNN の方が個人依存の影響が顕著な点である．他の手法は 5%程度の精度差なのに対し，DNN は約 15%の精度差がある．HC 特徴量を用いる場合，個人依存の影響を顕著に受ける可能性が示唆された．もう一つは，自己データを学習できる場合の DNN が非常に高い推定精度を達成することである．2-fold CV の結果で見ると，$SE\text{-}VGG \geq DNN > Li2018$ となり，DNN が Li2018 を上回るだけでなく SE-VGG に追従する推定精度となる．この原因の考察は容易ではないが，CNN により獲得した表現が，HC 特徴量と比べ個人に依存していないことに起因するものと考えている．HC 特徴量は個人依存の影響が顕著であるため，裏を返せば自己データを学習できることで特定個人の行動を識別するための決定境界を過学習気味に学習している可能性がある．

5. おわりに

本稿では，センサデータを用いた行動認識について幅広くサーベイし，行動認識で用いられている認識手法について体系的に取りまとめた．CNN を用いた表現学習モデルによる行動認識について十分な検討がなされていなかったことから，画像認識分野で研究が進んでいる CNN モデルをベースに，行動認識における有効性の検証実験を行った．行動認識では，子供と成人の歩行動作が同じでないように，個人依存による推定精度の低下が起こることが知られている．そこで，実験時には個人依存を踏まえた統一的な実験を実施し，各モデルの有効性を議論した．

行動認識のベンチマークデータセット HASC コーパスを用いて，HC 特徴量を用いた DNN，シンプルな CNN モデル，AlexNet，FCN，VGG，ResNet，SENet の推定精度検証を行った．その結果，訓練データに被験者を多く確保できる（多様性を担保できる）場合には，SE-VGG が最も高い精度を達成できた．一方で，訓練データの被験者数が極端に少ない場合は HC 特徴量による推定が有効な可能性も示唆した．今後は本稿をベンチマークとし，SE-VGG を基準にした改良モデルを模索することで，より高度な行動認識向けの表現学習モデルを検討していく予定である．

謝辞 本研究の一部はスズキ財団の科学技術研究助成によるものである．ここに謝意を表す．

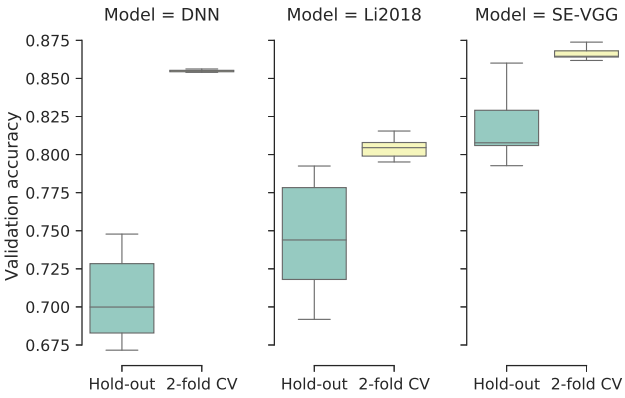

図 6 CNN モデルごとの個人依存度合いの比較
Fig. 6 Comparison of individual dependency degree for each CNN model.

参考文献

[1] Bao, L. and Intille, S. S.: Activity Recognition from User-Annotated Acceleration Data, *In Proceedings of the International Conference on Pervasive Computing (PerCom'04)* (2004).

[2] Kiani, K., Snijders, C. J. and S.Gelsema, E.: Recognition of daily life motor activity classes using an artificial neural network, *Archives of Physical Medicine and Rehabilitation*, Vol. 79, No. 2, pp. 147–154 (1998).

[3] Aminian, K., Robert, P., Buchser, E. E., Rutschmann, B., Hayoz, D. and Depairon, M.: Physical activity monitoring based on accelerometry: validation and comparison with video observation, *Medical & Biological Engineering & Computing*, Vol. 37, No. 3, pp. 304–308 (1999).

[4] Ravi, N., Dandekar, N., Mysore, P. and Littman, M. L.: Activity Recognition from Accelerometer Data, *In Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3*, AAAI Press, pp. 1541–1546 (2005).

[5] 村尾和哉，寺田　努：加速度センサの定常性判定による動作認識手法，情報処理学会論文誌，Vol. 52, No. 6, pp. 1968–1979 (2011).

[6] 大内一成，土井美和子：携帯電話搭載センサによるリアルタイム生活行動認識システム，情報処理学会論文誌，Vol. 53, No. 7, pp. 1675–1686 (2012).

[7] Kwapisz, J. R., Weiss, G. M. and Moore, S. A.: Activity Recognition Using Cell Phone Accelerometers, *SIGKDD Explor. Newsl.*, Vol. 12, No. 2, pp. 74–82 (online), DOI: 10.1145/1964897.1964918 (2011).

[8] Voicu, R.-A., Dobre, C., Bajenaru, L. and Ciobanu, R.-I.: Human Physical Activity Recognition Using Smartphone Sensors, *Sensors*, Vol. 19, No. 3 (online), DOI: 10.3390/s19030458 (2019).

[9] Shoaib, M., Bosch, S., Incel, O. D., Scholten, H. and Havinga, P. J. M.: Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors, *Sensors*, Vol. 16, No. 4 (online), DOI: 10.3390/s16040426 (2016).

[10] Xu, C., Pathak, P. H. and Mohapatra, P.: Finger-writing with Smartwatch: A Case for Finger and Hand Gesture Recognition Using Smartwatch, *In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pp. 9–14 (online), DOI: 10.1145/2699343.2699350 (2015).

[11] Bengio, Y., Courville, A. and Vincent, P.: Representation Learning: A Review and New Perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1798–1828 (2013).

[12] Frédéric Li, Kimiaki Shirahama, M. A. N. L. K. and Grzegorzek, M.: Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors, *Sensors*, Vol. 18, No. 679, pp. 1–22 (2019).

[13] Kawaguchi, N. and et al.: HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings, *In Proceedings of the Augmented Human International Conference (AH'11)* (2011).

[14] Alanezi, K. and Mishra, S.: Design, implementation and evaluation of a smartphone position discovery service for accurate context sensing, Vol. 44, pp. 307–323 (2015).

[15] Chen, L., Hoey, J., Nugent, C. D., Cook, D. J. and Yu, Z.: Sensor-Based Activity Recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 42, No. 6, pp. 790–808 (online), DOI: 10.1109/TSMCC.2012.2198883 (2012).

[16] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R. and Havinga, P.: Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey, *In Proceedings of the 23th International Conference on Architecture of Computing Systems 2010*, pp. 1–10 (2010).

[17] Lara, O. D. and Labrador, M. A.: A Survey on Human Activity Recognition using Wearable Sensors, *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 3, pp. 1192–1209 (2012).

[18] Altun, K. and Barshan, B.: Human Activity Recognition Using Inertial/Magnetic Sensor Units, *Human Behavior Understanding*, Vol. 6219, pp. 38–11 (2010).

[19] Yang, J.: Toward Physical Activity Diary: Motion Recognition Using Simple Acceleration Features with Mobile Phones, *In Proceedings of the 1st International Workshop on Interactive Multimedia for Consumer Electronics*, pp. 1–10 (2009).

[20] Breiman, L.: Random forests, *Machine Learning*, Vol. 45, pp. 5–32 (2001).

[21] Hinton, G. E. and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol. 313, No. 5786, pp. 504–507 (online), DOI: 10.1126/science.1127647 (2006).

[22] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pp. 1097–1105 (2012).

[23] Chen, Y. and Xue, Y.: A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer, *In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1488–1492 (online), DOI: 10.1109/SMC.2015.263 (2015).

[24] Gjoreski, H., Bizjak, J., Gjoreski, M. and Gams, M.: Comparing deep and classical machine learning methods for human activiy recognition using wrist accelerometer, *In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (ICJAI 2016)*, pp. 1–7 (2016).

[25] Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., Doppler, J., Holzmann, C., Kurz, M., Holl, G., Chavarriaga, R., Sagha, H., Bayati, H., Creatura, M. and d. R. Millán, J.: Collecting complex activity datasets in highly rich networked sensor environments, *In Proceedings of the 2010 Seventh International Conference on Networked Sensing Systems (INSS)*, pp. 233–240 (online), DOI: 10.1109/INSS.2010.5573462 (2010).

[26] Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S., Gerhard Tröster, J. d. R. M. and Roggen, D.: The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, *Pattern Recognition Letters*, Vol. 34, No. 15, pp. 2033–2042 (2013).

[27] Ha, S., Yun, J. and Choi, S.: Multi-modal Convolutional Neural Networks for Activity Recognition, *In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3017–3022 (online), DOI: 10.1109/SMC.2015.525 (2015).

[28] Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L. and Tröster, G.: Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection, *In Proceedings*

of the 5th European Conference on Wireless Sensor Networks, EWSN'08, pp. 17–33 (online), available from ⟨http://dl.acm.org/citation.cfm?id=1786014.1786017⟩ (2008).

[29] Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A. and Villalonga, C.: mHealthDroid: A Novel Framework for Agile Development of Mobile Health Applications, *In Proceedings of the Ambient Assisted Living and Daily Activities*, pp. 91–98 (2014).

[30] Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O., Zhu, J., Wu, P. and Zhang, J.: Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors, (online), DOI: 10.4108/icst.mobicase.2014.257786 (2014).

[31] Lockhart, J. W., Weiss, G. M., Xue, J. C., Gallagher, S. T., Grosner, A. B. and Pulickal, T. T.: Design Considerations for the WISDM Smart Phone-based Sensor Mining Architecture, *In Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, SensorKDD '11, pp. 25–33 (online), DOI: 10.1145/2003653.2003656 (2011).

[32] Hannink, J., Kautz, T., Pasluosta, C. F., Gaßmann, K., Klucken, J. and Eskofier, B. M.: Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks, *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 1, pp. 85–93 (online), DOI: 10.1109/JBHI.2016.2636456 (2017).

[33] Micucci, D., Mobilio, M. and Napoletano, P.: UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones, *Applied Sciences*, Vol. 7, No. 10 (2017).

[34] Yang, J., Nguyen, M. N., San, P. P., Li, X. L. and Krishnaswamy, S.: Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition, *In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI 2015, pp. 3995–4001 (2015).

[35] Yang, Z., Raymond, O. I., Zhang, C., Wan, Y. and Long, J.: DFTerNet: Towards 2-bit Dynamic Fusion Networks for Accurate Human Activity Recognition, *IEEE Access*, Vol. 6, pp. 56750–56764 (online), DOI: 10.1109/ACCESS.2018.2873315 (2018).

[36] Reiss, A. and Stricker, D.: Introducing a New Benchmarked Dataset for Activity Monitoring, *In Proceedings of the2012 16th International Symposium on Wearable Computers*, pp. 108–109 (online), DOI: 10.1109/ISWC.2012.13 (2012).

[37] Reiss, A. and Stricker, D.: Creating and Benchmarking a New Dataset for Physical Activity Monitoring, *In Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '12, pp. 40:1–40:8 (online), DOI: 10.1145/2413097.2413148 (2012).

[38] Xu, C., Chai, D., He, J., Zhang, X. and Duan, S.: InnoHAR: A Deep Neural Network for Complex Human Activity Recognition, *IEEE Access*, Vol. 7, pp. 9893–9902 (online), DOI: 10.1109/ACCESS.2018.2890675 (2019).

[39] Anguita, D., Ghio, A., Oneto, L., Parra, X. and Reyes-Ortiz, J. L.: A Public Domain Dataset for Human Activity Recognition Using Smartphones, *In Proceedings of the21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013*, pp. 437–442 (2013).

[40] Zhao, Y., Yang, R., Chevalier, G., Xu, X. and Zhang, Z.: Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors, *Mathematical Problems in Engineering*, Vol. 2018 (2018).

[41] Dong, M. and Han, J.: HAR-Net:Fusing Deep Representation and Hand-crafted Features for Human Activity Recognition (2018).

[42] Long, J., Sun, W., Yang, Z., Raymond, O. I. and Li, B.: Dual Residual Network for Accurate Human Activity Recognition (2019).

[43] Wang, J., Chen, Y., Hao, S., Peng, X. and Hu, L.: Deep learning for sensor-based activity recognition: A survey, *Pattern Recognition Letters*, Vol. 119, No. 1, pp. 3–11 (2019).

[44] Szegedy, C., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A.: Going deeper with convolutions, *In Proceedings of the2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9 (online), DOI: 10.1109/CVPR.2015.7298594 (2015).

[45] Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H. and Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, *In Proceedings of the-Conference on Empirical Methods in Natural Language (EMNLP 2014)*, pp. 1–9 (online), DOI: 10.1109/CVPR.2015.7298594 (2015).

[46] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).

[47] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (online), DOI: 10.1109/CVPR.2016.90 (2016).

[48] Glorot, X., Bordes, A. and Bengio, Y.: Deep Sparse Rectifier Neural Networks, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 15, pp. 315–323 (online), available from ⟨http://proceedings.mlr.press/v15/glorot11a.html⟩ (2011).

[49] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2014).

[50] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440 (online), DOI: 10.1109/CVPR.2015.7298965 (2015).

[51] Wang, Z., Yan, W. and Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline, *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1578–1585 (online), DOI: 10.1109/IJCNN.2017.7966039 (2017).

[52] Karim, F., Majumdar, S., Darabi, H. and Chen, S.: LSTM Fully Convolutional Networks for Time Series Classification, *IEEE Access*, Vol. 6, pp. 1662–1669 (online), DOI: 10.1109/ACCESS.2017.2779939 (2018).

[53] Hu, J., Shen, L. and Sun, G.: Squeeze-and-Excitation Networks, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

[54] Kingma, D. P. and Ba, J. L.: Adam: a Method for Stochastic Optimization, *International Conference on Learning Representations 2015*, pp. 1–15 (2015).