

# ウェアラブル心拍センサによる ワーク・エンゲイジメントの推定

原嶋 春輝<sup>1,a)</sup> 荒川 豊<sup>1,b)</sup> 石田 繁巳<sup>2,c)</sup> 中村 優吾<sup>1,d)</sup>

**概要:** 本研究は、ウェアラブルデバイスから得られるオフィスワーカーの日々の生体・活動状態からワーク・エンゲイジメント (WE: Work Engagement) を把握し、個人に適したサポートをすることを目的としている。本稿では、日本の一般企業5社に勤める60名のオフィスワーカーを対象に、ウェアラブルデバイスから2-3週間収集された日常的な心拍数データを使用した。また Utrecht Work Engagement Scale (UWES) と呼ばれる質問票を使用し、日々の WE スコアを計測した。これらのデータを用いて、機械学習手法による2種類の分析を行った。WE スコアの高低2群を対象とする、WE スコアの二値レベル推定では、Leave One Person Out 交差検証の結果、F1 値は 0.522 であった。一方、個人内の WE スコアにおける前日からの相対的变化を対象とする WE スコアの変動推定では、Leave One Person Out 交差検証の結果、F1 値は 0.663 であった。

## 1. はじめに

近年、心身の状態や周囲の環境に関するデータを利用して、労働者のパフォーマンスや快適性を向上させることに注目が集まっている。特に日本では、生産年齢人口の減少から、人手不足感による従業員の働きがいや意欲の低下が問題視されており、従業員が生き活きと働ける職場づくりが企業にとっての課題である。そのため、職場内の行動やコミュニケーションを計測したり [1], 姿勢を計測する [2,3] と行った研究が進められてきた。このようなフィジカルな状態も重要であるが、本研究では内面状態に焦点を当てる。コロナウイルス感染症に伴う、テレワークの増加により、これまで以上に働く人の心のケアが必要になるということが報告されており [4], 内面状態の把握は今後ますます重要性が高まると考えられる。

仕事に関連するポジティブで充実した心理状態を把握

する指標として、ワーク・エンゲイジメント (WE: Work Engagement) が挙げられる。Schaufeli らは、WE を「仕事に関連するポジティブで充実した心の状態であり、活力、熱意、没頭によって特徴づけられる」と定義しており、心身の疲労によって仕事への熱意を失った状態であるバーンアウトとは対の概念であると考えられている [5]。また、Harter らによると、従業員の WE と労働生産性には相関関係があることが報告されており、WE を高めることは従業員だけでなく組織にも大きなメリットがあると考えられている [6]。日本においては、モチベーションクラウド<sup>\*1</sup> や Wevox<sup>\*2</sup> などのツールを導入し、オンラインでのアンケートによって従業員の WE を管理する企業も増加している。

一般的に WE の把握には、Utrecht Work Engagement Scale (UWES) などの質問票によるアンケートが用いられる。しかし、このようなアンケートは定性的であり、主観的な評価による誤差が生じる可能性がある。また、月に一回程度の回答が一般的であり、時間や費用などの観点からも常時測定することが難しいという問題もある。

このような背景から、本研究では、手軽に利用できるウェアラブルデバイスを用いて労働者の日々の生体情報をモニタリングし、パフォーマンス向上や健康管理のサポートを行うことを目的とし、機械学習により、センサデータから WE を日常的に把握することを検討する。本研究では、WE を推定するための生体情報として、心拍を用い

<sup>1</sup> 九州大学大学院システム情報科学府/研究院 〒 819-0395 福岡市西区元岡 744

Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan

<sup>2</sup> 公立はこだて未来大学システム情報科学部 〒 041-8655 北海道函館市亀田中野町 116-2

School of Systems Information Science, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido, 041-8655 Japan

a) harashima.haruki.802@s.kyushu-u.ac.jp

b) arakawa@ait.kyushu-u.ac.jp

c) ish@fun.ac.jp

d) y-nakamura@ait.kyushu-u.ac.jp

<sup>\*1</sup> <https://www.motivation-cloud.com/>

<sup>\*2</sup> <https://get.wevox.io/>

表 1 センサを用いた質問票の推定研究

質問票	参照	センサ値
DAMS	Coutts et al. [8]	HRV
	Fukuda et al. [10]	sleep information
STAI	Coutts et al. [8]	HRV
	Mozos et al. [11]	HRV, electrodermal activity, microphone, acceleration
PSS	Coutts et al. [8]	HRV
	Sano et al. [12]	acceleration, screen time, GPS, skin conductance
PANAS (Positive and Negative Affect Schedule)	Muaremi et al. [13]	heart rate, audio, acceleration, GPS, contact frequency
SAM (Stress-Appraisal Measure)	Giakoumis et al. [14]	acceleration, HRV, dermal activity, video
WHOQOL (Quality of life for international comparison)	Amemori et al. [15]	acceleration, GPS, electrodermal activity, heart rate, skin temperature
OLBI (Oldenburg Burnout Inventory)	Garcia-Ceja et al. [16]	acceleration
UWES	-	-

る。心拍，特に心拍変動 (HRV : Heart Rate Variability) は自律神経活動の影響を受けていることから，メンタルヘルスに関する分析にも利用されている。例えば Huang らは，心電計から得られる HRV 指標を用いて，精神疲労状態の検出を試みている [7]。二値化した精神疲労状態の分類を行い，正解率は平均 75.5%であったことを報告した。また Coutts らは手首に装着したウェアラブルデバイスから HRV を取得し，高低 2 群に二値化した憂鬱な気分，前向きな気分，不安な気分などのメンタルヘルス状態の推定を行っている [8]。深層学習による推定の結果，高精度で分類可能であることが示された。しかし，彼らの実験は学生を対象としており，労働者の仕事に関するメンタルヘルス状態に関しては分析されていない。

また Baethge らは，公務員から日常的に取得した HRV データと，アンケートによって計測された WE の関係を分析した [9]。その結果，交感神経活性化と WE には関連性があると結論付けているが，心拍に関する特徴量による日々の WE の推定に関しては言及されていない。

本稿では，60 名のオフィスワーカーを対象に，ウェアラブルデバイスによって 2-3 週間収集された心拍データを使用した。また同時に実施された UWES による労働衛生アンケートの回答結果から，日々の WE スコアを算出した。これらのデータを用いて 2 種類の分析を行い，それぞれ二値分類機械学習モデルを構築した。

分析 1 では，関連研究における手法に従い，WE スコアを高低 2 群に二値化した，WE レベルの推定を行った。Leave One Person Out 交差検証の結果，F1 値は 0.522 であった。特に低い WE レベルの検出精度は 0.940 であり，WE が低い労働者を見逃す可能性は非常に低いと考えられる。

分析 2 では，各被験者の WE スコアが前日と比べて低下した場合，前日から変化なし・上昇した場合の 2 つのカテゴリに分類し，推定を行った。この推定手法は，日常的な心身のサポートにおける実用的なアプローチであると考えられる。Leave One Person Out 交差検証の結果，F1 値は 0.663 を示した。特に前日と比べて WE が低下しなかった場合の検出精度は 0.865 であった。

これらの推定結果は，それぞれ実用化に耐えられる精度には達していなかったが，これら 2 つのモデルを利点を活用することや，ウェアラブルデバイスから得られる心拍以外の情報も利用することによる発展が期待できる。

## 2. 関連研究

### 2.1 センサデータによる質問票の回答推定

メンタルヘルスの評価には，様々な質問票が用いられている。例えば Spielberger State-Trait Anxiety Inventory (STAI) は応用心理学研究において不安気分の測定に最も頻繁に用いられる。また，Perceived Stress Scale (PSS) は日常生活における慢性ストレスや将来への懸念から生じるストレス，周囲の状況によるストレスなど，自覚されるストレスレベルを測定する。一方，Depression and Anxiety Mood Scale (DAMS) は，抑うつ気分，肯定気分，不安気分の度合いを測定するために用いられる質問票である。

このような質問票に対する回答結果を，多様なセンサと機械学習などの手法を用いて推定する研究がされている。表 1 に，過去の研究で行われたメンタルヘルスの推定に関して，用いられた質問票とセンサ値について示す。メンタルヘルスの推定研究において様々な質問票が用いられているが，仕事に関連した心理状態である WE の推定は行われ

ていない。

また、表 1 から、メンタルヘルス状態を推定する研究では、心拍や HRV に関する特徴量がしばしば用いられていることが確認できる。

## 2.2 心拍変動指標によるメンタルヘルスの推定

HRV 指標は自律神経活動の影響を受けていることから、これまで心疾患や精神疾患の評価に用いられており、近年ではメンタルヘルス分野においても応用されている。

Huang らは、心電計を用いて精神疲労状態の検出可能性を調査した [7]。この論文では、35 名の学生を対象に筆記試験を実施した。筆記試験の前後において、独自に作成された 14 項目のアンケートによって精神疲労状態が計測された。また同時に心電計によって HRV 指標を 5 分間隔で収集している。時間領域及び周波数領域の 8 つの HRV 指標から、二値化された精神疲労状態を推定する機械学習モデルが構築され、5 分割交差検証による最高の平均正解率は 75.5% であることが報告された。

Huang らの方法のように、一般的に HRV 指標は心電計によって計測される心拍間隔 (RRI : R-R Interval) から算出される。しかし近年では、心電図に代わる方法も提案されている。例えば Coutts らは、652 名の学生が手首に装着したウェアラブルセンサから得られる時間領域及び周波数領域、その他の基本的な HRV 指標を用いて、アンケートから計測されるメンタルヘルス状態データを分析した [8]。メンタルヘルス状態を把握するための質問票には、PSS, STAI, DAMS が用いられた。統計的分析の結果、二値化されたメンタルヘルス状態間においていくつかの HRV 指標に有意な差が見られることから、HRV 指標はメンタルヘルスの診断に有効であることを主張している。また時系列解析などで用いられる深層リカレントニューラルネットワークである LSTM を用いて、各メンタルヘルス指標を高低 2 群に分類するモデルが構築された。約 2,000 件の 5 分間隔 HRV データセットおよび約 50 万件の 2 分間隔 HRV データセットを用いて、それぞれ最大 83% および 73% の分類精度が達成されたことを報告している。

これらの学生を対象とした実験の結果は、HRV がメンタルヘルス状態を把握することに有効であることを示唆しているが、労働者の仕事に関する心理状態の推定は行われていない。また精神状態やメンタルヘルス状態を二値化したラベルの推定が行われているが、日々の変動に関する推定は実施されていない。

## 2.3 心拍変動指標と WE の分析

Baethge らは、118 名の公務員を対象に 5 日間周波数領域の HRV 指標を計測し、交感神経活性化と WE の関係について分析している [9]。WE は、Utrecht Work Engagement Scale (UWES) に基づくアンケートによって取得された。

マルチレベル分析の結果、個人間における仕事、休暇中、睡眠中の交感神経活性化と WE にはそれぞれ正の相関関係があるという仮説が検証された。

Baethge らの研究では、HRV 指標と WE には関連性があることが示唆されているが、WE の推定に関する提案はされていない。一方、本研究では手首装着型のウェアラブルデバイスから得られる心拍・HRV 特徴量のみを用いて、日々の WE を推定可能であるかを検証する。

## 3. 提案手法

### 3.1 データ収集

本研究では、谷らのセンシングプロジェクトにおいて収集されたデータセットを使用した [17]。谷らは日本の 5 つの企業に勤めるオフィスワーカー 60 人を対象に 2-3 週間の日常生活における心身状態の計測を実施した。被験者には、性別や年齢などの事前アンケートへの回答が義務付けられている。また被験者は実験期間中、心拍数を取得可能な Fitbit Charge 3 \*<sup>3</sup> を 1 日中装着しながら生活し、毎朝 9 時にスマートフォンで WE に関するアンケートに回答した。収集されたアンケートは合計 569 件であり、Fitbit から取得された心拍数データは 13,424,969 レコードで 1.18GB であった。

### 3.2 アンケート回答の処理

オフィスワーカーの WE は、UWES に基づいた以下の質問によって測定している。

- 「今、仕事をしていると、活力がみなぎるように感じる」
- 「今、仕事に熱心である」
- 「今、私は仕事にのめり込んでいる」

被験者は 3 項目の質問に対して、それぞれ 0-6 点の点数で回答することを求められた。これらの合計点数を WE スコアとして今回の分析に用いた。この値は 0-18 の範囲を取り得る。また、未回答の項目を含むデータや、実験期間を通して WE スコアに変化がないデータを削除した。

分析 1 では、WE スコアをデータ全体の中央値を基準として、高低 2 群に分類し、ラベル付けを行った。分析 2 では、前日のワーク・エンゲイジメントスコアに対する翌日のスコアの変動によってラベル付けを行った。具体的には、「低下」、「変化なし・または上昇」の 2 つのラベルである。図 1 は、アンケートによって計測された被験者ごとの WE スコアの分布を示す。これは中央値やばらつきに大きな違いがあることを表している。一方で図 2 は、各被験者の日々の WE スコアの変動の分布を示している。中央値が 0 点 (前日から変化なし) 付近に集まっていることから、各被験者の WE スコアの低下数と上昇数には大きな偏りがない。

\*<sup>3</sup> <https://www.fitbit.com/global/jp/home>

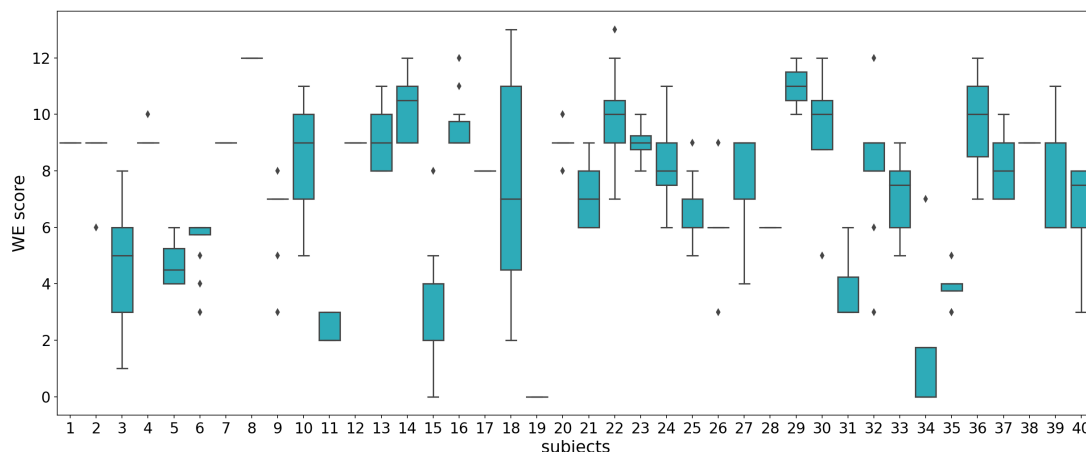


図 1 各被験者における WE スコアの分布

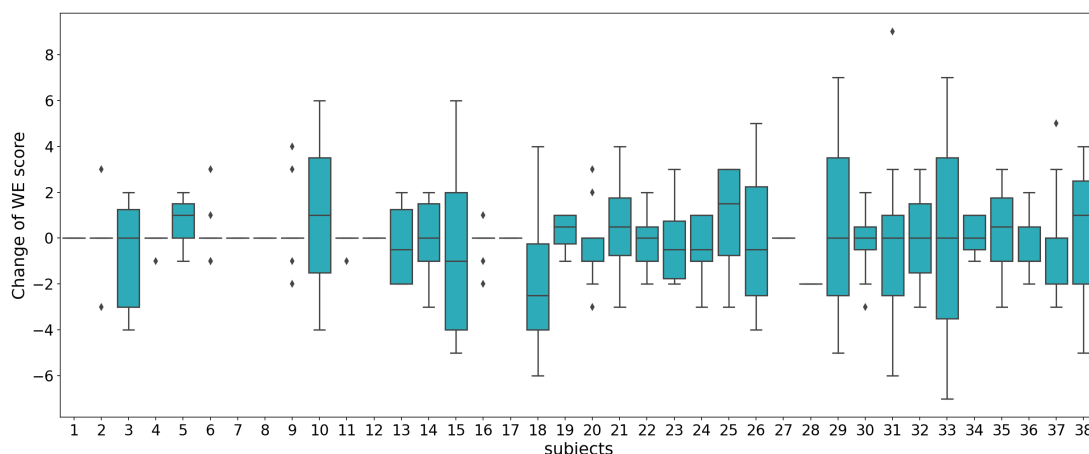


図 2 各被験者における WE スコア変動の分布

いことを示唆している。

### 3.3 心拍に関する特徴抽出

HRV 特徴量は、心拍間隔あるいは脈拍間隔である RRI から算出される。しかし、光電容積脈波法 (Photoplethysmography: PPG) が用いられている Fitbit Charge 3 では、心拍数がサンプリング周波数約 0.2Hz 記録されるだけである。そこで、福田らは、式 (1) によって心拍数から、RRI を算出している [18]。従って、本研究でも同様の方法で RRI の算出を行った。

$$RRI = \frac{60}{HeartRateData} \times 1000 \quad (1)$$

#### 3.3.1 予備調査

HRV 特徴量は一般的に時間領域特徴量と周波数領域特徴量の 2 種類が用いられる。しかし、心電計などの RRI を計測可能な機器と比較して、Fitbit の心拍数の計測はサンプリング周波数は非常に小さいことから、特に周波数領域特徴量が正確に算出できないことが懸念される。そこで予

備調査として、直接 RRI を取得可能なデバイスと Fitbit から算出した RRI について、周波数成分の比較を行った。

RRI を取得可能なデバイスとして、サンプリング周波数 1000Hz で計測を行い、心拍ごとに RRI データが記録されるウェアラブル心拍センサ WHS-1 を用いた\*4。WHS-1 及び Fitbit を装着した学生 1 名から、30 分間のデータ収集を 3 セット実施した。データ取得期間中の行動はオフィスワーカーを想定し、デスクワークと徒歩による移動のみに限定した。

周波数領域特徴量には低周波や高周波のパワースペクトル密度が用いられることから、それぞれのデバイスから得られる RRI のパワースペクトル密度を可視化することによって比較を行う。一般的な周波数領域 HRV 特徴量では、RRI のパワースペクトル密度において、0.04Hz 未満を超低周波成分 (VLF)、0.04–0.15Hz を低周波成分 (LF)、0.15–0.4Hz を高周波成分 (HF) として用いる。WHS-1 か

\*4 <https://www.uniontool.co.jp/product/sensor/whs1.html>

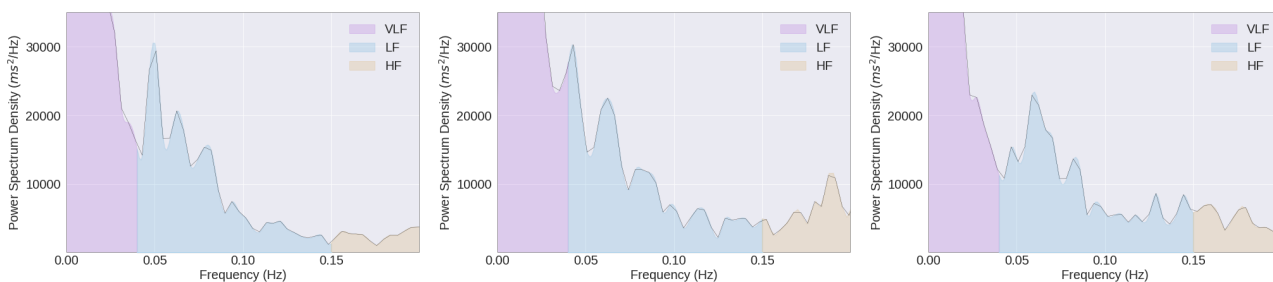


図 3 WHS-1 による RRI のパワースペクトル密度

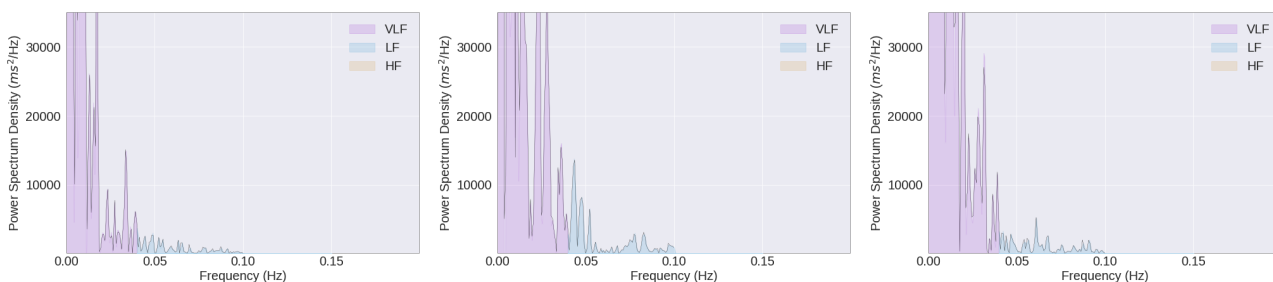


図 4 Fitbit による RRI のパワースペクトル密度

表 2 心拍データによる基本特微量

特微量の概要	
1	心拍数の平均値
2	心拍数の中央値
3	心拍数の標準偏差
4	心拍数の最大値
5	心拍数の最小値
6	RRI の平均値 (meanNN)
7	RRI の標準偏差 (SDNN)
8	隣接する RRI の差の二乗平均平方根 (RMSSD)
9	RRI の分散
10	ポアンカレプロットの縦軸方向標準偏差 (SD1)
11	ポアンカレプロットの横軸方向標準偏差 (SD2)

ら取得された RRI 及び Fitbit から算出した RRI のパワースペクトル密度を図 3 図 4 に示す。

図 3 は HRV 解析における一般的なパワースペクトル密度を示したものであるが、図 4 から分かるように、Fitbit による周波数領域特微量は高周波成分を全く表すことができない。また低周波成分においても大きな差異が確認できることから、Fitbit から算出された RRI の周波数領域 HRV 特微量は信頼性に欠けると考えられる。

この予備調査を踏まえ、本研究の分析では周波数領域の HRV 特微量は使用せず、時間領域特微量及び心拍数に関する特微量を用いることとする。

### 3.3.2 本研究で用いる特微量

本研究では、以下の 2 つの期間において、心拍数データから時間領域 HRV 特微量を含むいくつかの特微量を算出した。睡眠期間の識別には、Fitbit の独自のアルゴリズムにより記録された入眠及び起床時刻を使用した。更に期間 1 において、睡眠期間全体、入眠後 1 時間、起床前 1 時間

の 3 つの区分のデータを使用した。

期間 1：睡眠中

期間 2：アンケート直前の 1 時間 (午前 8-9 時)

また作成した基本的な 11 種類の特微量について表 2 に示す。ポアンカレプロット (またはローレンツプロット) は各 RRI と直前の RRI の関係を散布図として表したものである。この散布図は一般的に楕円形を示し、これらの縦軸及び横軸方向の標準偏差 (SD1, SD2) は自律神経活動の評価に用いられている [19,20]。

分析 1 では、2 つの期間において算出された 11 種類の基本的な特微量に加え、各特微量における期間 1 と期間 2 の差分及び比率を新たな特微量として追加した。分析 2 では、分析 1 で用いる特微量に加え、それらの前日との比率を新たな特微量として追加した。また各分析の特微量について、2 つの異なるラベル間で統計的有意差検定を行うことで、2 群間の差を表す特微量の調査及び特微量の選択を試みた。

特微量を作成する前段階において、2 つのデータ取得期間において長時間一定の心拍数が記録されたデータは異常値とみなし、削除した。アンケート回答結果と心拍数データを紐付け、それぞれの分析において最終的に使用可能なデータセットは 317 件、275 件であった。

また、それぞれの分析に使用可能なデータ件数は被験者間で異なっており、分析 1 における被験者間のデータ件数について、最小値は 1 件、中央値は 7 件、最大値は 15 件であった。分析 2 における被験者間のデータ件数の最小値は 1 件、中央値は 8 件、最大値は 15 件であった。

### 3.4 モデル構築

2つのいずれの分析においても、分類モデルには Light Gradient Boosting Machine (LightGBM) を使用した。LightGBM は勾配ブースティング決定木の1つであり、精度及び計算速度ともに優れており、近年よく利用されている機械学習アルゴリズムである。分析1ではWEスコアの二値レベル(高い, 低い)の分類, 分析2ではWEスコアの変化(低下, 不変または上昇)の二値分類として、モデルの学習を行った。

モデルの評価には、Leave One Person Out (LOPO) 交差検証を用いた。LOPO 交差検証とは、1人の被験者のデータを評価用データセット、残りの被験者のデータを学習データセットとしてモデルを構築する。これを全ての被験者ごとに繰り返すことで、モデルの被験者間の汎化性能を評価する方法である。評価指標として、正解率、適合率、再現率、F1値を用いた。

## 4. 結果と考察

### 4.1 分析1

分析1では、WEスコアを高低に二値化したWEレベルの推定を行った。LOPO 交差検証による評価の結果、正解率は0.576、F1値は0.522であった。この時の混同行列を図5に、LOPO 交差検証によって被験者ごとに評価された正解率、適合率、再現率、F1値の箱ひげ図を図6に示す。また、統計的有意差検定の結果、高低2群のWEレベル間において有意な差のある特徴量は確認できなかった。従って、分析1では特徴量の選択は実施していない。

表3はLightGBMモデルでの分類における平均的な特徴重要度の上位10個を表す。睡眠時及び、睡眠時に対するアンケート前の比率に関する特徴量が上位を占めている。睡眠時は日中に比べて心拍数が安定することから、各被験者の基準値になっていると考えられる。この基準値と仕事に取り組んでいる時の値の比率はWEの推定を行うために有効な特徴量であると思われる。比率に関する特徴量の中でも、最も重要度の高いRMSSDは、副交感神経と強い関係があると考えられており、WEレベルの差は副交感神経活動による影響を受けている可能性も示唆される。

図5から、低いWEレベルの検出は高精度であった一方、高いWEレベルの検出は優れていないことが分かる。また図6に示すように、被験者間における各指標のばらつきは大きく、正解率及びF1値における四分位範囲はそれぞれは0.60、0.54である。このような結果の原因として、2点のことが考えられる。まず1点目はデータ件数の少ない被験者も推定対象としていることである。本分析で使用したデータのうち、被験者間におけるデータ件数は最小で1件であった。このようなデータ件数の極めて少ない被験者を対象としたLOPO 交差検証では、評価値は極端な値を取りやすくなると考えられる。2点目は、個人間でWEス

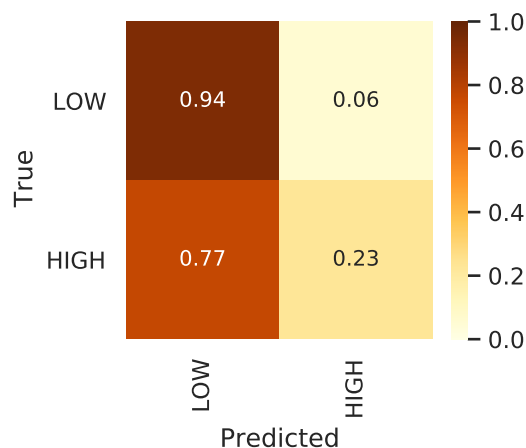


図5 分析1における混同行列

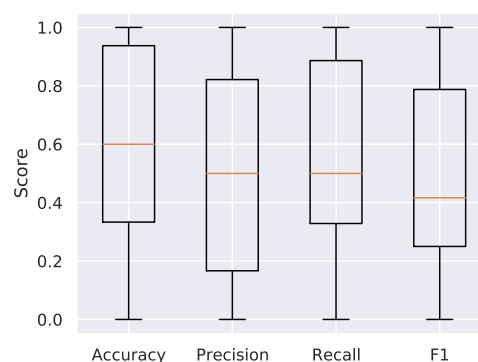


図6 分析1における被験者ごとの評価結果

コアの基準が異なる可能性を考慮していないことが挙げられる。図1から分かるように、被験者ごとにWEスコアの中央値やばらつきが大きく異なっている。しかし本分析では、WEスコアをデータ全体の中央値である8点で二値化を行っていることから、被験者内におけるラベルの偏りが生じてしまい、各被験者内のWEレベルの差を表現できていない。従って、被験者ごとの精度を向上させるためには、各被験者のアンケートに対する主観的基準の違いを考慮する必要があると考えられる。具体的な対策として、ユーザーごとのデータからWEレベルの基準を設定し、パーソナライズする方法が挙げられる。そのためには、個人ごとにある程度のデータ数を用意しなければならないことが課題である。

### 4.2 分析2

分析2では、個人内におけるWEスコアの日々の相対的変動(前日と比べて低下または不変・上昇)の推定を行った。LOPO 交差検証による評価の結果、正解率は0.709、F1値は0.597であった。

また、WEスコアの変動を示す2群間(前日と比べて低

表 3 分析 1 の推定における特徴重要度

特徴量	注釈	重要度
心拍数の最大値	期間 1	0.044
RMSSD	2 つの期間の比率	0.033
心拍数の最小値	2 つの期間の比率	0.033
心拍数の平均値	期間 2	0.029
心拍数の標準偏差	2 つの期間の比率	0.029
心拍数の最小値	起床前期間	0.029
SD2	入眠後期間	0.029
心拍数の平均値	2 つの期間の比率	0.028
心拍数の中央値	2 つの期間の比率	0.027
心拍数の最大値	2 つの期間の比率	0.024

表 4 分析 2 における前日との比率に関する特徴量

特徴量	注釈	p 値 <sup>1</sup>	重要度
心拍数の標準偏差	期間 1	0.047	0.097
心拍数の平均値	期間 2	0.007	0.068
心拍数の中央値	期間 2	0.011	0.051
心拍数の標準偏差	期間 2	0.033	0.151
meanNN	期間 2	0.013	0.130
心拍数の平均値	2 つの期間の比率	0.015	0.095
心拍数の最大値	2 つの期間の比率	0.049	0.062
心拍数の標準偏差	2 つの期間の比率	0.006	0.079
meanNN	2 つの期間の比率	0.034	0.096
SDNN	2 つの期間の比率	0.029	0.000
RRI の分散	2 つの期間の比率	0.029	0.072
SD2	2 つの期間の比率	0.026	0.101

<sup>1</sup> マン・ホイットニーの U 検定 (有意水準 0.05)

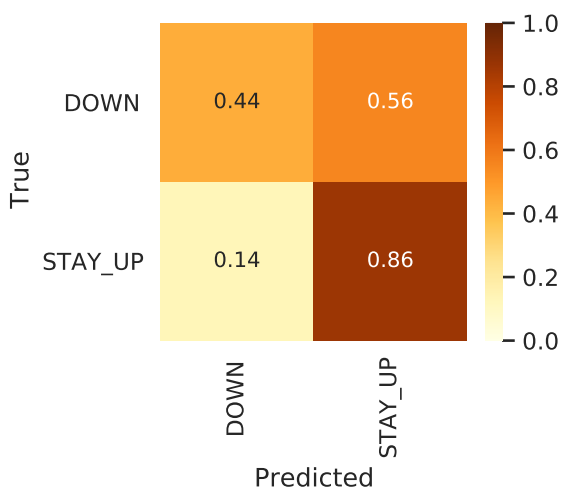


図 7 分析 2 における混同行列

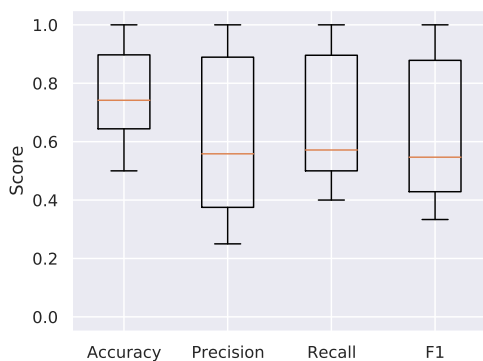


図 8 分析 2 における被験者ごとの評価結果

下, 不変または上昇) において, 全ての特徴量について統計的有意差検定を行ったところ, 前日との比率に関する 12 個の特徴量に有意な差が確認された. この 12 個の特徴量のみを使用して LOPO 交差検証を行った場合, 正解率は 0.727, F1 値は 0.663 となった. 表 4 に, 有意差のある特徴量と, LightGBM の分類における重要度を示す. 有意差のあった, 前日との比率に関する特徴量において, 特に期間 2 における心拍数の標準偏差及び meanNN, は高い重要度を示した. またこの時の混同行列及び評価結果の箱ひげ

図を図 7, 図 8 に示す.

混同行列から, WE スコアが低下した場合の検出に比べて, 変化しないまたは上昇した場合の検出に優れていることが示唆される. また, 被験者ごとの評価における正解率及び F1 値の四分位範囲それぞれは 0.25, 0.45 である. 分析 1 の評価結果と比較して, 各評価指標の四分位範囲のばらつきが小さいことから, 絶対的な WE レベルを推定する方法より, WE スコアの相対的な変動を推定する方法は, 汎化性能が高いアプローチであると考えられる.

## 5. おわりに

本研究では, 日本の一般企業 5 社に勤務する 60 名のオフィスワーカーを対象に, ウェアラブルデバイスによって 2-3 週間収集された心拍数データを使用した. また同時に実施された UWES による労働衛生アンケートの回答結果から, 日々の WE スコアを算出した. 分析 1 では, WE スコアの二値レベル (高い, 低い) の分類を行った. Leave One Person Out 交差検証の結果, F1 値は 0.522 を示した. 分析 2 では, WE の日々の変動 (前日と比べて低下または不変・上昇) の分類を行った. Leave One Person Out 交差検証の結果, F1 値は 0.663 であった.

分析 1・分析 2 のいずれにおいてもオフィスワーカーの WE をモニタリングするための実用化に耐え得る精度には達していなかった. しかし分析 1 における WE レベルの推定に比べ, 分析 2 の WE スコアの変動の推定はより高い精度で推定可能であり, データや特徴量に多様性を持たせることによる更なる精度の向上が期待される. また, それぞれの推定モデルの利点を組み合わせることや, ウェアラブルデバイスから得られる睡眠や歩数などの情報を利用することによる発展の余地は大きいと考えられる. 今回の 2 種類の分析ではそれぞれ二値分類による推定を行ったが, ラベルを細分化し, より有用的な推定を行うことも課題である. 更に実用化に向けてさらに多くの被験者で実験を行い, 汎化性能について検証する必要がある.

謝辞 本研究の一部は、JSPS 科研費 JP18H03233, および東北大学電気通信研究所の共同研究プロジェクトプログラムの支援を受けて実施したものである。

#### 参考文献

- [1] Ara, K., Kanehira, N., Olguin, D. O., Waber, B. N., Kim, T., Mohan, A., Gloor, P., Laubacher, R., Oster, D., Pentland, A. S. et al.: Sensible organizations: Changing our businesses and work styles through sensor data, *Journal of Information Processing*, Vol. 16, pp. 1–12 (2008).
- [2] Otoda, Y., Mizumoto, T., Arakawa, Y., Nakajima, C., Kohana, M., Uenishi, M. and Yasumoto, K.: Census: Continuous posture sensing chair for office workers, *2018 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, pp. 1–2 (2018).
- [3] Nakamura, Y., Matsuda, Y., Arakawa, Y. and Yasumoto, K.: WaistBelt X: A Belt-Type Wearable Device with Sensing and Intervention Toward Health Behavior Change, *Sensors*, Vol. 19, No. 20, p. 4600 (2019).
- [4] Shimazu, A., Nakata, A., Nagata, T., Arakawa, Y., Kuroda, S., Inamizu, N. and Yamamoto, I.: Psychosocial impact of COVID-19 for general workers, *Journal of occupational health*, Vol. 62, No. 1, p. e12132 (2020).
- [5] Schaufeli, W. B., Martinez, I. M., Pinto, A. M., Salanova, M. and Bakker, A. B.: Burnout and engagement in university students: A cross-national study, *Journal of cross-cultural psychology*, Vol. 33, No. 5, pp. 464–481 (2002).
- [6] James K. Harter, Frank L. Schmidt, S. A. S. K. P.: THE RELATIONSHIP BETWEEN ENGAGEMENT AT WORK AND ORGANIZATIONAL OUTCOMES, <https://employeeengagement.com/wp-content/uploads/2013/04/2012-Q12-Meta-Analysis-Research-Paper.pdf> (2012).
- [7] Huang, S., Li, J., Zhang, P. and Zhang, W.: Detection of mental fatigue state with wearable ECG devices, *International journal of medical informatics*, Vol. 119, pp. 39–46 (2018).
- [8] Coutts, L. V., Plans, D., Brown, A. W. and Collo-mosse, J.: Deep learning with wearable based heart rate variability for prediction of mental and general health, *Journal of Biomedical Informatics*, Vol. 112, p. 103610 (2020).
- [9] Baethge, A., Junker, N. M. and Rigotti, T.: Does work engagement physiologically deplete? Results from a daily diary study, *Work & Stress*, pp. 1–18 (2020).
- [10] Fukuda, S., Matsuda, Y., Tani, Y., Arakawa, Y. and Yasumoto, K.: Predicting Depression and Anxiety Mood by Wrist-Worn Sleep Sensor, *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, pp. 1–6 (2020).
- [11] Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R. and Ferrandez, J. M.: Stress detection using wearable physiological and sociometric sensors, *International journal of neural systems*, Vol. 27, No. 02, p. 1650041 (2017).
- [12] Sano, A. and Picard, R. W.: Stress recognition using wearable sensors and mobile phones, *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, IEEE, pp. 671–676 (2013).
- [13] Muaremi, A., Arnrich, B. and Tröster, G.: Towards measuring stress with smartphones and wearable devices during workday and sleep, *BioNanoScience*, Vol. 3, No. 2, pp. 172–183 (2013).
- [14] Giakoumis, D., Drosou, A., Cipresso, P., Tzovaras, D., Hassapis, G., Gaggioli, A. and Riva, G.: Using activity-related behavioural features towards more effective automatic stress detection, *PLoS one*, Vol. 7, No. 9, p. e43571 (2012).
- [15] Chishu Amemori, Teruhiro Mizumoto, Y. A. and Yasumoto, K.: Simplified Measurement Method for HRQOL based on WHOQOL-BREF by Smart Devices, *Multimedia, Distributed, Cooperative, and Mobile Symposium*, Vol. 2017, pp. 880–887 (2017).
- [16] Garcia-Ceja, E., Osmani, V. and Mayora, O.: Automatic stress detection in working environments from smartphones’ accelerometer data: a first step, *IEEE journal of biomedical and health informatics*, Vol. 20, No. 4, pp. 1053–1060 (2015).
- [17] Tani, Y., Fukuda, S., Matsuda, Y., Inoue, S. and Arakawa, Y.: WorkerSense: Mobile Sensing Platform for Collecting Physiological, Mental, and Environmental State of Office Workers, *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, pp. 1–6 (2020).
- [18] 福田修之, 松田裕貴, 荒川豊, 安本慶一ほか: 健常者のうつ兆候検知に向けた睡眠状況と労働衛生指標の統計分析, 研究報告モバイルコンピューティングとパーベシブシステム (MBL), Vol. 2020, No. 11, pp. 1–8 (2020).
- [19] Mourot, L., Bouhaddi, M., Perrey, S., Rouillon, J.-D. and Regnard, J.: Quantitative Poincare plot analysis of heart rate variability: effect of endurance training, *European journal of applied physiology*, Vol. 91, No. 1, pp. 79–87 (2004).
- [20] Jeppesen, J., Beniczky, S., Johansen, P., Sidenius, P. and Fuglsang-Frederiksen, A.: Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot, *Seizure*, Vol. 24, pp. 1–7 (2015).