

令和 4 年度
修士学位論文

サーモグラフィより観測した
呼吸パターンからのストレス推定への
深層学習の応用

An Application of Deep Learning to
Stress Estimation Based on Breathing Patterns
through Thermography

1255103 笥 拓也

指導教員 岩田 誠

2023 年 2 月 28 日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要 旨

サーモグラフィより観測した 呼吸パターンからのストレス推定への 深層学習の応用

笥 拓也

近年，心拍数や心拍変動，脳波等の生体信号を用いたストレス推定法が提案されている [1]．多くの手法では，接触型センサによる計測が必要なため日常生活の中で簡易なストレス推定が困難である．一方，非接触型センサを用いた手法も研究されており，赤外線カメラより観測した呼吸パターン (RVS) に基づくストレス推定法 [2] が提案された．

本研究では，文献 [2] と同一データセットを用いて多層 2D-CNN と ImageNet に基づく転移学習によるストレス推定を試みた．2 クラス分類では転移学習により精度が 10% 程度向上した一方，3 クラス分類ではいずれも向上せず，特に転移学習では，“Low-level-stress” の F 値が数% となった．これは，RVS と主観ストレス強度の個人差が大きく，各ストレスのラベルに偏りがあることに起因すると考えられる．また，RVS は “Low-level-stress” 固有の特徴量が少なく，多層の畳み込みを有する転移学習では十分に学習できなかったと考えられる．そこで，本稿では，データセットとモデル性能の改善点に着目して “Low-level-stress” の F 値向上が可能な深層学習の応用法について検討する．

本研究では，新たな心理実験にて収集したデータセットに対する提案手法によって “Low-level-stress” の F 値が最大で 45% 程度向上した．しかし，精度は向上しなかったため，今後は複数の生体信号を用いた深層学習の手法についての検討が必要である．

キーワード サーモグラフィ，RVS(Respiration Variability Spectrogram)，主観ストレス強度，ストレス推定，2D-CNN，転移学習

Abstract

An Application of Deep Learning to Stress Estimation Based on Breathing Patterns through Thermography

Takuya Kakehi

In recent years, stress estimation methods using biological signals such as heart rate, heart rate variability, and electroencephalography have been proposed[1]. Most of those methods require measurement with contact sensors so that it is difficult to casually estimate stress in daily life. On the other hand, methods using non-contact sensors have also been studied. A stress estimation method based on respiration patterns (RVS) observed through an infrared camera was proposed[2].

In this study, stress estimation was attempted with multi-layer 2D-CNN and transfer learning based on ImageNet using the dataset same as in the literature[2]. The result of “Low-level-stress” F value was a few percent. This may be attributed to the large individual differences in RVS, subjective stress intensity, and the bias in the labels of each stress. In addition, RVS has few “Low-level-stress” specific features and could not be sufficiently learned by transfer learning with multi-layer convolution. Therefore, this research examined the application of deep learning to improve the F value of “Low-level-stress” by focusing on the improvement of the dataset and model performance. In this study, the proposed method improved the F value of “Low-level-stress” by about 45% at maximum on a dataset collected in our owned psychological experiment. However, the accuracy did not improve more, this indicates that future studies on deep learning

methods using multiple biological signals are needed.

key words Thermography, RVS(Respiration Variability Spectrogram), Subjective stress intensity, Stress estimation, 2D-CNN, Transfer learning

目次

第 1 章	序章	1
第 2 章	RVS に基づくストレス推定とその課題	6
2.1	緒言	6
2.2	RVS に基づくストレス推定法 [2]	6
2.2.1	心理実験によるデータ収集	7
2.2.2	RVS 生成や主観ストレス強度に関する前処理	8
2.2.3	深層学習による RVS に基づくストレス推定	11
2.3	RVS に基づくストレス推定の追検証	11
2.4	本研究の方針	14
2.5	結言	15
第 3 章	RVS に着目した精度向上の改善策	17
3.1	緒言	17
3.2	本研究における実験	17
3.3	心理実験によるデータ収集	17
3.4	前処理およびデータセット構築	20
3.5	深層学習における特徴抽出能力の改善	20
3.6	結言	21
第 4 章	評価と考察	22
4.1	緒言	22
4.2	参加者データの解析	22
4.2.1	主観ストレス評価の解析	22
4.2.2	課題正答率および反応速度の解析	25

目次

4.3	ストレスの推定精度	26
4.3.1	各手法単独の有効性についての検証	27
4.3.2	複数の手法を組み合わせた有効性についての検証	42
4.4	考察	43
4.5	結言	46
第 5 章	結論	47
	謝辞	50
	参考文献	51

目次

1.1	ストレスを感じている労働者の推移 ([4] を基に作成)	2
1.2	ストレスの発生過程	3
2.1	RVS に基づくストレス推定の手続き	7
2.2	文献 [2] における実験手続き	8
2.3	個人毎に正規化された主観ストレス強度のスコア [2]	9
2.4	データセット [2] における各ストレスに対する RVS	10
2.5	データセット [2] における個人毎のラベルの分布	10
2.6	2 層 2D-CNN[2] による各ストレスクラスの分類結果	12
3.1	本研究における実験手続き	18
3.2	本研究の心理実験にて使用した Digital-VAS	19
3.3	本研究の心理実験における実験環境	19
3.4	Transfer learning の手順	21
4.1	Test 直後の主観ストレス評価のみを対象とした場合における各課題難易 度間の平均値の差についての検定結果	23
4.2	1 セット目の Relax 直後と各課題の easy 実施後の主観ストレス評価につ いての検定結果	24
4.3	Relax 直後と Test 直後の主観ストレス評価の差分を対象とした場合にお ける各課題難易度間の平均値の差についての検定結果	25
4.4	課題難易度別正答率の検定結果	26
4.5	課題難易度別反応速度の検定結果	27
4.6	先行研究データ [2] に Balanced を適用したラベルの分布	29
4.7	本実験データに Balanced を適用したラベルの分布	35

図目次

4.8 Separated 適用時の 3 クラス分類時のラベル分布	39
---	----

表目次

2.1	データセット [2] の 2 クラス分類時における精度と F 値	14
2.2	データセット [2] の 3 クラス分類時における精度と F 値	14
4.1	Augmented の適用の有無の比較	28
4.2	水平方向と垂直方向の比較	28
4.3	先行研究データ [2] を対象とした Over 適用時の 2 クラス分類の精度と F 値	30
4.4	先行研究データ [2] を対象とした Over 適用時の 3 クラス分類の精度と F 値	30
4.5	先行研究データ [2] を対象とした Down 適用時の 2 クラス分類の精度と F 値	31
4.6	先行研究データ [2] を対象とした Down 適用時の 3 クラス分類の精度と F 値	31
4.7	Baseline の適用の有無の比較	33
4.8	Cut 適用時の 2 クラス分類の精度と F 値	34
4.9	Cut 適用時の 3 クラス分類の精度と F 値	34
4.10	本実験データを対象とした Over 適用時の 2 クラス分類の精度と F 値 . .	36
4.11	本実験データを対象とした Over 適用時の 3 クラス分類の精度と F 値 . .	36
4.12	本実験データを対象とした Down 適用時の 2 クラス分類の精度と F 値 . .	37
4.13	本実験データを対象とした Down 適用時の 3 クラス分類の精度と F 値 . .	38
4.14	Separated 適用時の 3 クラス分類の精度と F 値	39
4.15	Transfer learning 適用時の 2 クラス分類の精度と F 値	40
4.16	Transfer learning 適用時の 3 クラス分類の精度と F 値	40
4.17	各手法単独の有効性についての結果	41
4.18	本実験データに Base+Cut+Bal.+Xfer を組み合わせた 3 クラス分類時の 精度と F 値の比較	42

表目次

4.19 本実験データに Base+Bal. を組み合わせた 3 クラス分類時の精度と F 値 の比較	43
4.20 感情に対する課題難易度間の検定結果 (有意水準 5%)	44
4.21 3 つの既存モデルによる感情の分類精度 (%)	45

第 1 章

序章

現代社会は、経済的豊かさや科学技術の発展によって、より快適で利便性に満ち溢れている反面、「ストレス社会」と言われている [3]。特に、労働者を取り巻く環境は少子高齢化や国際競争の激化、人員削減による負担増加等により大きく変化している。それに伴い、仕事上でストレスを抱える労働者の割合が増加し、ストレスの原因が年々変化している [4]。厚生労働省が実施している労働安全衛生調査（実態調査）の結果によると、仕事上で感じるストレスの内容は特に「仕事の量」、「仕事の質」、「仕事の失敗、責任の発生等」が多いと報告されている [5]。また、図 1.1 に示すよう仕事上でストレスを抱える労働者の割合は、半数を超える数字で推移しており、1990 年代後半からは約 6 割となっている。ストレスの蓄積によって、自律神経失調症や突発性難聴、適応障害等を引き起こす可能性があり、精神面だけでなく、身体面の不調の原因、さらに、大病の誘引や発病にもつながり得る [6]。そのため、今日のストレス社会において、自身のストレスの状態を正確に理解することや個人を取り巻く周囲の支援を充実させることは健康の維持や病気の予防に重要な意味を持つ [7]。

そもそも、ストレスという言葉の定義は非常に難しいものであり、悲しみや苦しみといった人々にとってマイナスな状態を意味する言葉として漠然と使われている。ストレスの語源は、工学・物理学の用語で、外から力を加えられた時に生じる歪みを表現したものとなっている。Selye は、その言葉を医学・生理学の用語として生体内に生じた歪みの状態を表現する抽象的な言葉として用いた。その生体側に生じる歪みが「ストレス (ストレス反応)」であり、歪みを生じさせる外部刺激は「ストレッサー」と呼ばれる。スト

仕事上でストレスを感じている労働者の割合

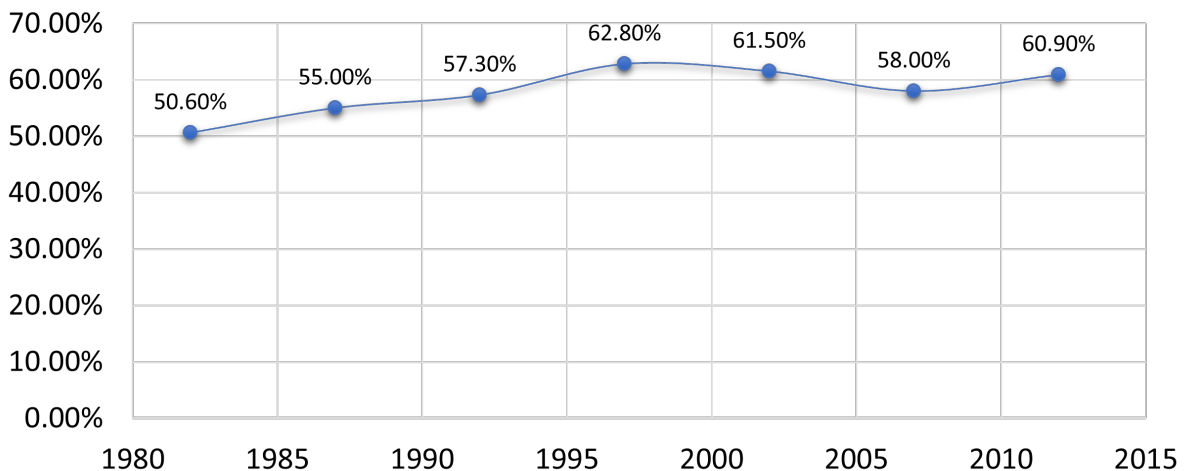


図 1.1: ストレスを感じている労働者の推移 ([4] を基に作成)

レッサーには 4 つの種類があり、騒音や寒冷等の物理的ストレスと栄養素や薬品等の化学的ストレス、細菌・ウイルス感染等の生物的ストレス、そして、不安や緊張等の心理的ストレスに分けられる [8]。これら、ストレスによってストレスは心理面、生理面、行動面の反応として表出される。特に、生理面の反応とは、血圧や心拍、呼吸、脳波といった生体信号のことであり、行動面の反応とは、表情、姿勢、刺激に対する反応時間等を指す。仮に複数人に対して同じストレスが与えられても個人によってストレス反応は異なり、ある人は脅威と感じ、またある人は機会と捉える。Selye は、これをストレスの非特異性として強調した [9]。これに関して、Lazarus と Folkman はストレスの発生の過程として、「ストレス」、「評価」、「ストレス反応」の 3 段階に分けられることを提唱した [10]。図 1.2 に示すよう、ストレスの発生過程ではまず、一次評価として、ストレスに対してそれが自分にとってどの程度脅威となり得るのか主観的な基準で判断する。次に二次評価では、ストレスを対処できるかどうか推測する。この際、ストレスに対して到底対処できないと認識することで、その人にとって大きなストレスとなり得る。このような評価過程を経てストレス反応が表出されるため、ストレス反応には個人差が生じる。なお、各ストレス反応に対するデータの収集方法は異なり、心理面の反応は、個人の主観評価による報告で測定され、生理面の反応の

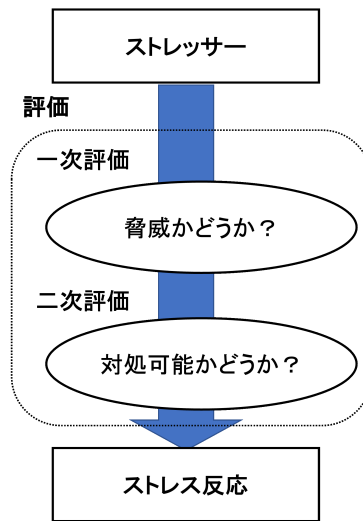


図 1.2: ストレスの発生過程

収集には専用のセンサや環境を必要とする。ただ、ストレス反応のデータ収集を目的とした実験を行った場合、心理面の反応に対する主観評価による測定では、実験参加者が意図的に虚偽の報告をすることや実験者が意図していない感情を感じる可能性がある。また、行動面の反応に関しても、ストレスを感じているのに反して、笑顔や嬉しそうな表情を作り出すために自身を意識的に操作することがしばしばある。その一方で、心理面や行動面の反応とは異なり、人間の根本にあたる生理面の反応は、実験参加者が意識的に操作できるものではないことから、ストレスに関する研究では、生理面の反応を識別指標として用いることが多い [8]。

近年では、ストレスに関してさまざまな分野において幅広く研究されている。心理学、医学・生理学の分野では、ストレスに対する心拍数や心拍変動、呼吸数、脳波といった様々な生理面の反応の理解を目指している。それと対照的に、工学の分野では生理面や行動面の反応に基づいたストレスの自動検出や日常生活の中でより簡易的にストレスを測定する方法等、現実世界への応用に重点が置かれている。Healey と Picard は、自動車を運転中のドライバーから心電図と筋電図、ガルバニック皮膚反応、そして呼吸パターンを収集し、最適な特徴量を用いて k 近傍法によってストレス強度を 4 段階に分類した [1]。Rimini-Doering らは、ドライビングシミュレータ時の眠気やストレス状態における

心電図、脳波、皮膚温度、呼吸周波数の他複数の生理面の反応や頭の動き、開眼といった行動面の反応の振る舞いについて研究している [11]. これらの研究では、実験参加者から生理学的指標を収集するために十分に統制された実験環境にて複数種類の接触型センサを使用する必要がある。しかし、接触型センサは装着によって拘束感が生じ、また、大掛かりな装置を必要とするため、持ち運びが難しく、データ測定時の環境が限られる等、制約が多い。これらは参加者にとって、身体的にも精神的にも大きな負担となり、本来実験者が意図していない生理学的反応やストレス等のデータが測定される可能性がある。このことから、接触型のセンサを用いた実験では、日常生活において自然に生じるストレスに対する生理学的指標の測定および推定が難しいのではないかという見方もある。その一方で、非接触型のセンサを用いたストレス研究も行われている。片岡らは、鼻と額の表面温度とストレスには高い相関があることを報告しており、非接触型の皮膚温度測定システムを開発している [12]. また、Engert らは、赤外線カメラによって記録した主要な顔面部の温度 (鼻先、額、口腔周囲、眼窩周囲) において鼻先と口腔周囲の温度低下が精神的なストレスの生理的指標になり得ることを明らかにした [13]. これらは一例ではあるが、非接触型センサを用いて顔面部の温度変化とストレスや感情の関係性について研究が進められている。通常、実験参加者毎に置かれている社会的背景や健康状態は異なっているため、実験を通して測定された生理面の反応は、ストレスと直接的に結びついていると考えるのが難しいと言われている。実際、生理面の反応一つひとつがストレスの有無や強度を識別する上でどのように関与しているのか明白ではないことについて言及されている [8]. しかし、2017 年に Cho らによって行われた研究では、赤外線カメラで撮影された鼻孔領域サーモグラフィから観測した呼吸パターン (RVS: Respiration Variability Spectrogram) のみから、ストレス強度を推定する手法が提案された [2]. 文献 [2] において、2 クラス分類時のストレス強度は、“None-stress” と “Stress” とされ、3 クラス分類時は、“None-stress”, “Low-level-stress”, “High-level-stress” とされており、2 層の 2D-CNN を用いた分類精度は 2 クラスで 84.6%, 3 クラスで 56.5%と報告されている [2]. この結果より、非侵襲的かつ簡易なストレス推定の実現可能性が示唆された。

しかし、2 クラス分類時の “None-stress” と 3 クラス分類時の “Low-level-stress” の精度はチャンスレベル付近となっており、十分に高くはない。そこで、本研究では、文献 [2] の著者から同一のデータセットを提供してもらい、多層 2D-CNN と ImageNet に基づく転移学習によるストレス推定を試みた。その結果、2 クラス分類時の精度は転移学習によって 10% 程度向上したが、3 クラス分類時はいずれも低下した。また、転移学習では 3 クラス分類時における “Low-level-stress” の F 値が数% となった。これは、大きくデータセットとモデル性能に起因する課題と考えられるため、本稿では、深層学習の枠組みにて改善策について検討を行う。

本稿の第 2 章では、文献 [2] の RVS に基づくストレス推定法の手続きについて述べる。その後、2 クラス分類時の “None-stress” と 3 クラス分類時の “Low-level-stress” の分類精度が低い原因を調査するために実施した追検証の結果から予測される課題に触れる。そして、3 クラス分類時の “Low-level-stress” の F 値向上を目的とした本研究における深層学習の改善方針について検討する。

第 3 章では、RVS に基づくストレス推定の 3 つの手続きである、心理実験によるデータ収集と RVS 生成や主観ストレス強度の正規化等の前処理、そして、深層学習によるストレス推定、それぞれに対して改善策の提案を行う。

第 4 章では、3 クラス分類時の “Low-level-stress” の F 値向上を目的とした改善手法の評価を行う。評価に先立ち、まず、心理実験にて収集した参加者データの解析を行い、主観ストレス強度と実施した課題の相関を明らかにする。その後、先行研究データ [2] と本実験により収集したデータを対象に、初めに、提案手法単独の有効性について検証し、その後、有効性の高い組み合わせにて評価を行う。そして、第 5 章では、本研究の総括、および、更なる精度向上に向けた課題考察と今後の展望を述べる。

第 2 章

RVS に基づくストレス推定とその課題

2.1 緒言

本章では，文献 [2] の RVS に基づくストレス推定法の手続きと深層学習による分類結果について述べる．その後，2 クラス分類時の “None-stress” と 3 クラス分類時の “Low-level-stress” の精度がチャンスレベル付近と十分に高くなかった原因を調査するために実施した追検証の結果に触れる．そして，最後に，RVS に基づくストレス推定の精度向上を目的とした本研究における改善策の検討方針を述べる．

2.2 RVS に基づくストレス推定法 [2]

2017 年に Cho らは，低価格な赤外線カメラ (FLIR One) から観測した呼吸パターンの RVS に基づくストレス推定法を提案した [2]．この手法では，RVS に基づくストレス推定に深層学習が用いられており，計算コストの小さい 2 層の 2D-CNN が採用された．それによる分類精度は，2 クラス分類時では 84.6%，3 クラス分類時では 56.6%と報告され，日常生活の中で非侵襲的かつ簡易にストレス推定する実現可能性が示唆された．

RVS に基づくストレス推定の手続きを図 2.1 に示す．本節では，図 2.1-1 の心理実験によるデータ収集から図 2.1-3 の深層学習を用いたストレス推定までの手続きについてそれぞれ解説する．

2.2 RVS に基づくストレス推定法 [2]

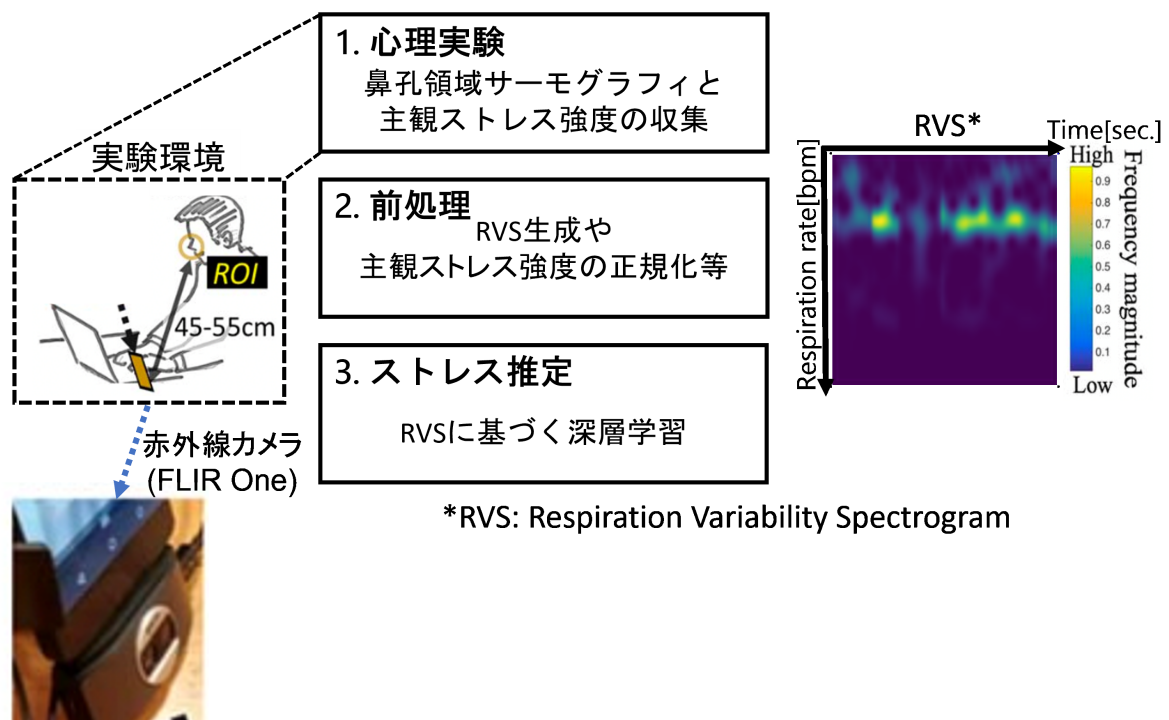


図 2.1: RVS に基づくストレス推定の手続き

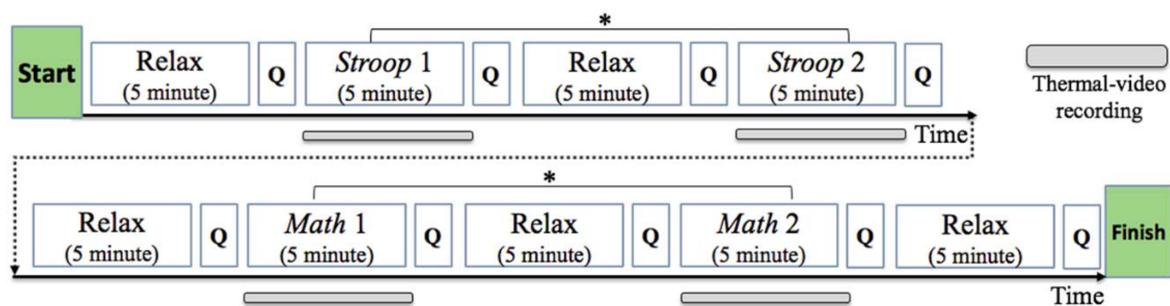
2.2.1 心理実験によるデータ収集

心理実験は、図 2.2 の手続きに基づいて、健常な成人 8 名 (男性 5 名, 女性 3 名) を対象に室温が管理されていない静寂な環境にて実施された。なお、実験参加者の平均年齢は $30.75(\pm 10.22)$ 歳だった。

実験では、まず初めに、図 2.2 の “Relax” に示す 5 分間のリラックス期間を過ごし、Relax 終了後に参加者は現在のストレス強度について質問された (i.e. Q)。この際、10cm の Visual Analog Scale (VAS) が用いられ、Q はその後の課題 (i.e. Stroop または Math) の前後にて実施された。VAS の左端、すなわち 0cm は not at all, 10cm は very much とされていた。ストレス強度の主観評価終了後、5 分間の課題を実施し、この時間のみ赤外線カメラによって鼻孔領域サーモグラフィを撮影した。赤外線カメラにはアンドロイドスマートフォンに取り付けられた FLIR One が使用され、カメラから鼻孔領域までの距離は 45cm~55cm に設定された。課題は、それぞれ easy と difficult の 2 段階の難易度のストロープ課題 (Stroop) と計算課題 (Math) で、それぞれの課題期間においてカウ

2.2 RVS に基づくストレス推定法 [2]

ンターバランスをとって実施された。Math の 1 問当たりの問題提示時間は 7.5 秒間で、Stroop は 1.5 秒間とされていた。また、Math では、回答に対して音声フィードバックが用いられ、Stroop では、“Correct” と “Incorrect” という文字のフィードバックが用いられた。



* ラテン方格によりeasyとdifficultでカウンターバランスをとって実施

図 2.2: 文献 [2] における実験手続き

2.2.2 RVS 生成や主観ストレス強度に関する前処理

心理実験にて収集されたデータは、図 2.1-2 の前処理によって、鼻孔領域サーモグラフィからは RVS が、主観ストレス強度からは RVS に対する教師データとなるラベルが生成された。まず、前者に関して、呼吸追跡アルゴリズム [14] を使用して鼻孔領域サーモグラフィから 1 次元の呼吸信号を復元する。その後、20 秒間の呼吸信号を 1 秒間隔でオーバーラップして短時間フーリエ変換を適用することで、呼吸変動といった動的な特徴量を含むスペクトログラムが取得される。これを 2 次元画像として個別に切り出すことで、RVS が生成される。RVS は図 2.1 に示すように、横軸に時間 [sec.]、縦軸に呼吸数 [bpm] を、色情報としてパワースペクトル密度を持つ。次に、後者に関して、主観ストレス強度は実験参加者間において個人差が生じるため、個人毎に正規化を行う。正規化された主観ストレス強度のスコアを図 2.3 に示すように、それぞれの課題では参加者に対して誘発させるストレススコアが異なることが確認される。正規化された主観ストレス強度はその後、k-means 法によってクラス数に応じたラベルに変換される。

2.2 RVS に基づくストレス推定法 [2]

2 クラス分類の場合, “None-stress” と “Stress”, 3 クラス分類の場合, “None-stress”, “Low-level-stress”, “High-level-stress” がラベルとなる. また, 双方のクラスのラベルは数値として扱われ, 2 クラス分類時は (“None-stress”, “Stress”) = (0, 1) とされ, 3 クラス分類時は (“None-stress”, “Low-level-stress”, “High-level-stress”) = (0, 1, 2) とされる.

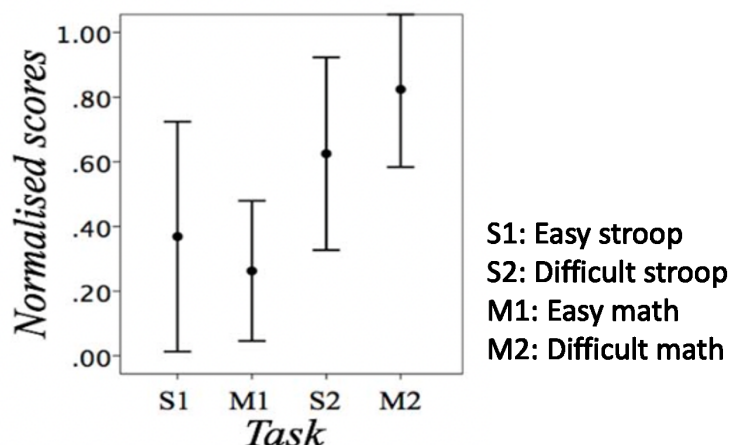


図 2.3: 個人毎に正規化された主観ストレス強度のスコア [2]

これらの前処理によって得られた 3936 画像の RVS とそれに対する主観ストレス強度のラベルからデータセットが構築されている. 図 2.4 にデータセットの一部を示す. この図から, 各ストレス強度に対する RVS の特徴が異なることが確認される. また, 図 2.5 に 2 クラス分類時と 3 クラス分類時における個人毎のラベルの分布を示す. これらから確認されるように, 双方のクラスにおいて個人毎に保有するラベル数は異なり, 特に, 2 クラス分類時のラベルに関して, p3, p5, p6 は “Stress” のラベルしか保有しておらず, “None-stress” が 1305 画像に対して “Stress” は 2631 画像と全体の 2/3 を占める (図 2.5-a). また, 3 クラスに関しても 3 つ全てのラベルを保有する参加者が p2 と p7 のみであり, さらに, p1 は “None-stress” のラベルが極端に多く, 個人毎に偏りが大きいことが確認される (図 2.5-b).

2.2 RVS に基づくストレス推定法 [2]

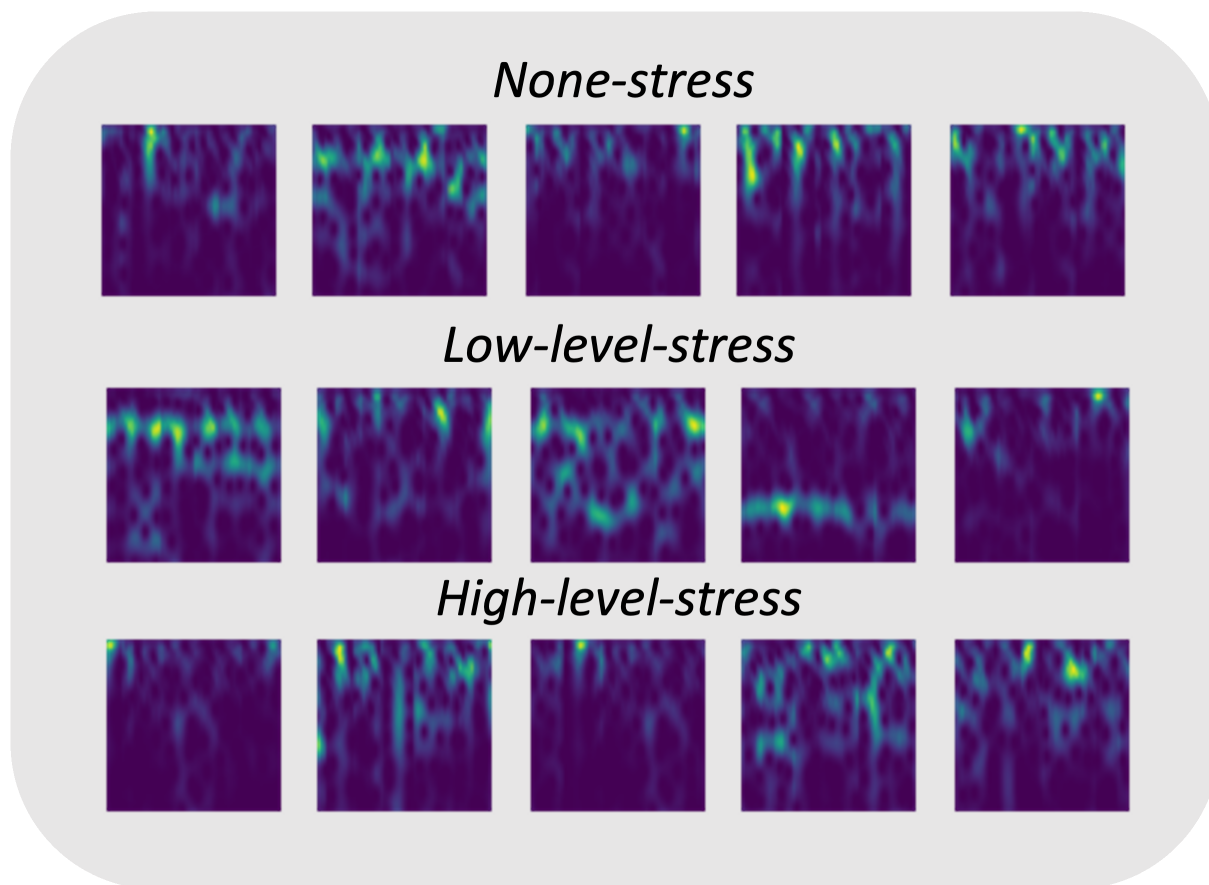


図 2.4: データセット [2] における各ストレスに対する RVS

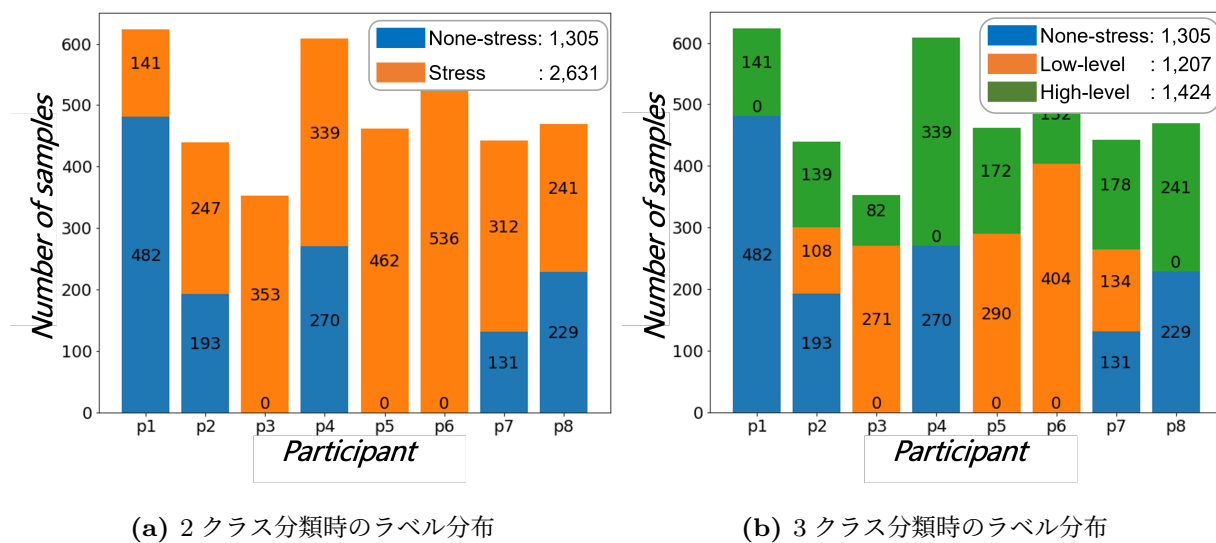


図 2.5: データセット [2] における個人毎のラベルの分布

2.3 RVS に基づくストレス推定の追検証

2.2.3 深層学習による RVS に基づくストレス推定

図 2.1-3 の深層学習による RVS に基づくストレス推定では、深層学習モデルとして 2 層 2D-CNN[2] が使用された。ネットワークの構成は、2D-Convolution(kernel = 5×5) と MaxPooling(2,2) を交互に 2 回繰り返す、その後、Fully connected を経て、Softmax によりクラス数に応じた出力がなされる。なお、学習時のハイパーパラメータは、エポックサイズ 300、バッチサイズ 50、学習率 0.5 で、入力画像サイズ 28×28 、最適化アルゴリズムには SGD が使用された。また、評価方法には 8-fold Leave-one-subject-out(Loso) 交差検証法が用いられた。

2 層 2D-CNN[2] によるストレスの推定精度は、2 クラス分類時で 84.6%、3 クラス分類時で 56.5%と報告されている。また、図 2.6 の混同行列にて 2 クラス分類時と 3 クラス分類時における双方の詳細な分類結果を示す。それぞれの混同行列における Actual Class は、実験参加者が実際に認知したストレスクラスを表しており、Predicted Class は 2 層 2D-CNN[2] が RVS から推定したストレスクラスである。2 クラス分類時の “Stress” は 94.1%で、3 クラス分類時の “High-level-stress” は 72.8%とそれぞれのクラスにおいて最も高精度で分類できている。その一方、2 クラス分類時の “None-stress” は 54.6%で、3 クラス分類時の “Low-level-stress” は 38.5%とチャンスレベル付近の値となっており、十分に高い精度とは言えない。

2.3 RVS に基づくストレス推定の追検証

本研究では、2 クラス分類時の “None-stress” と 3 クラス分類時の “Low-level-stress” の分類精度が十分に高くなかった原因を調査するために、文献 [2] の著者から同一のデータセットを提供して頂き、2 層 2D-CNN[2] に加え、多層 2D-CNN(3 層、5 層) と ImageNet にて事前学習された 3 つの既存モデル (ResNet-18[15], MobileNet-v3[16], EfficientNet-b0[17]) による転移学習によってストレス推定を試みた。その際、評価方法は文献 [2] と同様に Loso 交差検証法とした。

2.3 RVS に基づくストレス推定の追検証

		Predicted Class	
		No-stress	Stress
Actual Class	No-stress	713 (54.6%)	592 (45.4%)
	Stress	128 (05.9%)	2503 (94.1%)

(a) 2 クラスの混同行列

		Predicted Class		
		No-stress	Low-level	High-level
Actual Class	No-stress	731 (56.0%)	286 (21.9%)	288 (22.1%)
	Low-level stress	282 (23.4%)	465 (38.5%)	460 (38.1%)
	High-level stress	0 (0.00%)	387 (27.2%)	1037 (72.8%)

(b) 3 クラスの混同行列

図 2.6: 2 層 2D-CNN[2] による各ストレスクラスのカテゴリ分け結果

新たに使用した多層 2D-CNN の構成と既存モデルによる転移学習時の設定とハイパーパラメータを以下に示す。

- 3 層 2D-CNN

特徴抽出器は、 $\{2D-Conv.(kernel = n \times n, filter = m) + Relu + MaxPooling(2,2)\}$ を 1 ブロックとして、3 ブロックの縦列接続の構成とした。kernel サイズ n は、1 ブロック目から順に 5, 5, 3 で、filter サイズ m は 2 クラス分類時では 9, 18, 24 として、3 クラス分類時では 12, 24, 48 に設定した。分類器については 2 層 2D-CNN と同一にした。

ハイパーパラメータに関して、エポックサイズは 150 としてそれ以外は先行研究 [2] と同じ設定とした。

- 5 層 2D-CNN

特徴抽出器は、 $\{2D-Conv.(kernel = n \times n, filter = m) + Relu\}$ の Block1 と $\{2D-Conv.(kernel = n \times n, filter = m) + Relu + MaxPooling(2,2)\}$ の Block2 を Block1-Block1-Block2-Block1-Block1 と縦列接続の構成とした。kernel サイズ n は、順に 5, 5, 3, 5, 5 で、filter サイズ m は、順に 9, 18, 24, 32, 48 と 2 クラス分類と 3

2.3 RVS に基づくストレス推定の追検証

クラス分類時において同じ設定とした。また、分類器については 2 層 2D-CNN と同一にした。

ハイパーパラメータに関しては、エポックサイズ 150, 学習率 0.05 としてそれ以外は先行研究 [2] と同じ設定とした。

- 既存モデルによる転移学習

全てのモデルにおいて分類器のみ今回のストレスクラスに応じた出力数に変更し、学習の際は ImageNet で事前学習された重みを基に先行研究データ [2] 用に新たに学習させた。その際、エポックサイズは 20 として学習率は 0.0001 に設定した。

2 クラス分類時の結果を表 2.1 に示す。論文では、2 層 2D-CNN[2] の分類精度が 84.6% と報告されていたのに対し、追検証の結果は 70.8% だった。また、その他の 2D-CNN では精度が 66.8~68.0% と下回る結果となった。その一方、転移学習では、いずれも精度が向上し、特に MobileNet-v3[16] では 82.7% となった。F 値に関してはいずれも “None-stress” より “Stress” の値が高い結果となった。

3 クラス分類時の結果を表 2.2 に示す。2 層 2D-CNN[2] の分類精度は論文の結果と類似する 55.8% であった。一方、その他いずれのモデルにおいても、精度が向上せず、10~14% 低下した。また、F 値に関しては、3 層 2D-CNN を除く全てのモデルにおいて “Low-level-stress” は “None-stress” と “High-level-stress” より低く、特に転移学習を適用した際は、“Low-level-stress” の F 値が 10% 未満と極端に低い値となった。

以上の結果より、転移学習は 2 クラス分類時において精度と F 値の向上に有効性があると考えられる。しかし、3 クラス分類時では、“Low-level-stress” の F 値が極端に低下し、数% となったことから、“None-stress” と “High-level-stress” に比べて RVS に基づくストレス推定が困難だと言える。これは、データセットの質とモデル性能が不十分であったことに起因する可能性がある。

2.4 本研究の方針

表 2.1: データセット [2] の 2 クラス分類時における精度と F 値

		精度	F 値	
			None	Stress
2D-CNN	2 層 2D-CNN[2]	70.81	44.89	80.15
	3 層 2D-CNN	68.01	20.16	80.00
	5 層 2D-CNN	66.79	40.29	77.00
転移学習	ResNet18[15]	71.95	40.65	81.64
	MobileNet-v3[16]	82.72	71.36	87.63
	EfficientNet-b0[17]	80.41	67.48	85.98

表 2.2: データセット [2] の 3 クラス分類時における精度と F 値

		精度	F 値		
			None	Low	High
2D-CNN	2 層 2D-CNN[2]	55.79	57.39	43.89	63.96
	3 層 2D-CNN	45.75	34.22	45.57	51.29
	5 層 2D-CNN	55.49	58.73	37.59	65.89
転移学習	ResNet18[15]	41.34	49.49	5.37	60.86
	MobileNet-v3[16]	43.80	66.67	9.00	50.17
	EfficientNet-b0[17]	44.99	63.23	7.52	54.57

2.4 本研究の方針

2.3 節にて実施した文献 [2] と同一のデータセットに対する追検証では、多層 2D-CNN と転移学習を用いた場合における 3 クラス分類時の精度が改善されておらず、特に、転移学習では、“Low-level-stress” の F 値が数%であることが課題と明らかになった。これはデータセットの質とモデル性能に起因する可能性がある。まず、データセットに関し

2.5 結言

て、生体信号である RVS と主観ストレス強度は個人差が大きく、加えて、8 人の実験参加者毎に 2 クラス分類時と 3 クラス分類時における各ストレスクラスのラベル数に偏りがあるため全てのストレス強度を均等に学習できないと考えられる。また、モデル性能に関して、RVS は “None-stress” や “High-level-stress” に比べて “Low-level-stress” 固有の特徴を含んでいないため、多層の畳み込みでは十分に学習できていないと考えられる。実際、3 クラス分類時において ImageNet に基づく転移学習では、2D-CNN に比べて “Low-level-stress” の F 値が “None-stress” と “High-level-stress” よりも低い。その一方、2 クラス分類時は “None-stress” と “Stress” の F 値が共に高い値であることから、ImageNet に基づく転移学習のように RVS の高次の特徴を学習するにつれ、ストレスクラスが “None-stress” と “High-level-stress” に 2 分されていることが考えられる。つまり、RVS は “Low-level-stress” 固有の特徴をあまり保有しておらず、分類精度が十分に高くないと考えられる。

そのため、RVS に基づくストレス推定までの 3 つの流れそれぞれに対し、改善策を講じる必要がある。そこで、まず初めに、条件を統制した心理実験にてデータを収集し、次に、個人差を緩和できる前処理およびラベルの均等化に着目して新たなデータセットを構築する必要がある。また、これらに加え、データ拡張および事前学習タスクの改善による転移学習の適用することで更なる特徴抽出能力の向上が期待される。

これらを踏まえ、本研究では、データセットとモデル性能の 2 つ側面から RVS に基づくストレス推定の精度向上に向けた深層学習の応用法について検討することで 3 クラス分類時における “Low-level-stress” の F 値向上を試みる。

2.5 結言

本章では、非接触型センサから観測した RVS に基づくストレス推定法 [2] の手続きについて述べた。また、2 層の 2D-CNN[2] による 2 クラス分類時の “None-stress” と 3 クラス分類時の “Low-level-stress” の精度が十分に高くない原因調査のために実施した多

2.5 結言

層 2D-CNN と転移学習によるストレス推定の結果に触れた。その結果より、3 クラス分類時における精度が改善されていない原因を考察し、‘Low-level-stress’ の F 値向上に対する改善点についてデータセットの質とモデル性能の観点から述べた。そして、最後に、RVS に基づくストレス推定の精度向上に向けた本稿における深層学習の応用手法の検討方針について述べた。

第 3 章

RVS に着目した精度向上の改善策

3.1 緒言

本章では、3 クラス分類時における “Low-level-stress” の F 値向上に向け、RVS に基づくストレス推定までの 3 つの流れ、それぞれに対してデータセットの質とモデル性能の改善点に着目した提案手法について述べる。

3.2 本研究における実験

2.4 節の本研究の方針にて述べたように、3 クラス分類時における “Low-level-stress” の F 値の向上には RVS に基づくストレス推定までの 3 つの流れそれぞれに対してデータセットとモデル性能の改善が必要と考えられる。そこで、本稿では、それらの改善点に対して、条件を統制した心理実験によるデータ収集および個人差を緩和できる前処理、そして、深層学習における特徴抽出能力の向上に向けた改善手法について提案する。

3.3 心理実験によるデータ収集

本研究における実験参加者は、健常な大学生 11 名 (男性 7 名, 女性 4 名) であり、平均年齢は 19.09(± 0.30) 歳であった。

実験の手続きは、文献 [2] を参考に 5 分間のリラックス期間 (Relax) と Relax 直後のストレスに関する主観評価 (Q), 5 分間の課題期間 (Test), そして、Test 直後の Q を 1 セットとした 4 セット試行 (およそ 60 分強) で構成した。図 3.1 に本研究にて実施し

3.3 心理実験によるデータ収集



図 3.1: 本研究における実験手続き

た実験手続きを示す。

- Relax

安静状態を維持するための 5 分間のリラックス期間を過ごしてもらった。この時間は、スマートフォンの閲覧や飲食、その場から移動することは厳禁として、リラックス効果があると考えられる動画を鑑賞してもらった。

- Q

ストレスに関する主観評価では、図 3.2 の Digital-VAS を用いて画面中央のスライダーをマウスで左右に動かして回答してもらった。なお、1 セット目の Relax 直後の Q を主観ストレス評価の基準とし、それを基に Test 後の Q の判断するように指示した。

- Test

- 5 分間の課題では、文献 [2] と同様にそれぞれ 2 段階の難易度 (easy と difficult) の計算課題 (Math) とストロープ課題 (Stroop) を各セットにおいてランダムに出題した。各課題の問題提示時間および正誤フィードバックの方法についても同様に設定した。
- 本研究では、個人差の抑制および各ストレスクラスのラベルの均等化に向け、各課題の difficult の難易度を上げることに加え、鼻孔領域サーモグラフィ撮影時に顎台で顔を固定した。

Math の easy は $\{5000 \rightarrow 1\}$ で、difficult は、 $\{5000 \rightarrow \text{random_int}(1,50)\}$ とした。実験参加者には、提示された問題に対してキーボードから答えとなる数

3.3 心理実験によるデータ収集



図 3.2: 本研究の心理実験にて使用した Digital-VAS

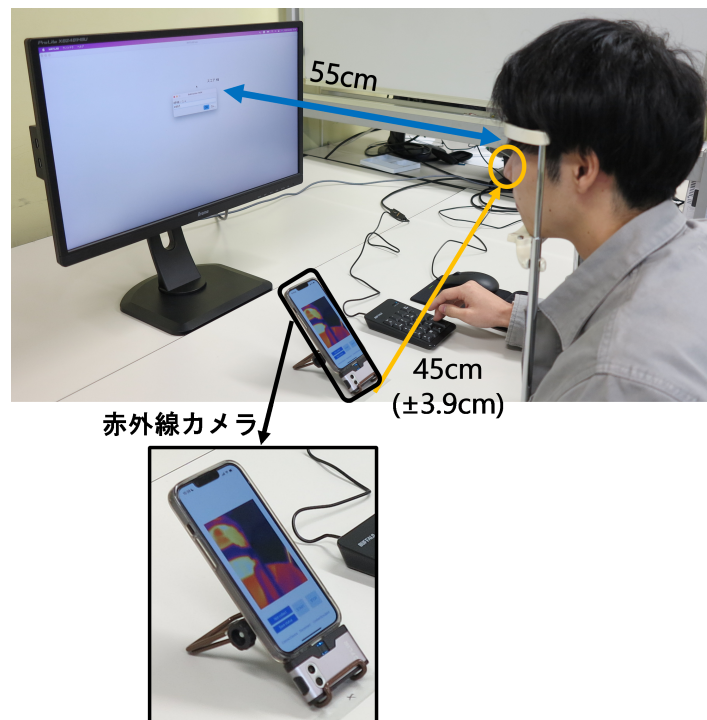


図 3.3: 本研究の心理実験における実験環境

値を入力させた。また、Stroop の easy は文字の意味と文字の色を一致させ、difficult では、文字の色と画面に表示する文字サイズをランダムに変更した。実験参加者には、提示された問題に対してマウスで答えとなる色を指定されたボタンをクリックして回答させた。

なお、本実験にて実施した課題プログラムは MATLAB2022a により作成した。

- サーモグラフィ撮影時には赤外線カメラに FLIR ONE Gen3 for iOS を使用した。本研究における実験環境を図 3.3 に示すように、顎台使用時のディスプレイ距離は 55cm で固定し、さらに、カメラ距離は 45cm(±3.9cm) に設定した。

3.4 前処理およびデータセット構築

3.4 前処理およびデータセット構築

個人差の抑制に向け、心理実験にて収集したデータに対して、以下の 1 と 2 の前処理を施し、データセットの均等化に関しては 3 の手法を適用した。さらに、4 の手法によって 3 クラス分類時における “Low-level-stress” の F 値の改善を試みた。

1. 主観ストレス強度のベースライン化による正規化 (**Baseline: Base**)

主観ストレス評価には個人差があるため、Test 直後から Relax 直後の差分をラベル化する。

2. 不安定な呼吸期間のデータの削除 (**Cut**)

課題開始直後は、緊張等により呼吸が安定していない可能性があるため、開始 30 秒間を削除して RVS を生成する。

3. 不均等なデータセットの均等化 (**Balanced: Bal.**)

2 通りの方法でデータセットの均等化を行う。

- 個人毎に少数派のラベルを多数派に合わせてランダムにオーバーサンプリング (Oversampling: Over)
- 個人毎に多数派のラベルを少数派に合わせてランダムにダウンサンプリング (Down-sampling: Down)

上記のいずれの方法においても 2 クラス分類時および 3 クラス分類時における各ストレスクラスのラベル全てを保有する参加者のデータのみ学習とテストに使用する。

4. データセットの課題種類別分割 (**Separated: Sep.**)

Math と Stroop では、認知されるストレス強度が異なると考えられるため、より効果的に弱いストレスを誘発している課題データセットを使用することで、3 クラス分類時における精度および各ストレスの F 値が向上すると期待される。

3.5 深層学習における特徴抽出能力の改善

深層学習における特徴抽出能力の改善として以下の 1 と 2 手法を適用した。

3.6 結言

1. データ拡張 (**Augmented: Aug.**)

水平・垂直方向の反転を $1/2$ の確率で実施した。

2. 事前学習タスクの改善 (**Transfer learning: Xfer**)

図 3.4 に示すように、まず初めに、ImageNet で事前学習された既存モデルを先行研究のデータセット [2] 用にファインチューニングし、最もロスが減少した際の重みを使用して、本実験で収集したデータセットの学習とテストに使用する。

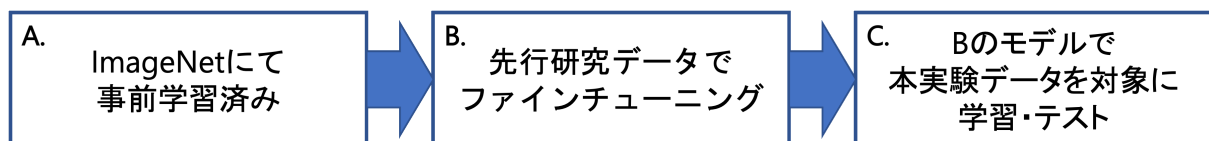


図 3.4: Transfer learning の手順

3.6 結言

本章では、3 クラス分類時における “Low-level-stress” の F 値向上に向け、RVS に基づくストレス推定までの流れに対する改善手法について述べた。まず、心理実験では、個人差の抑制と各ストレスクラスのラベルの均等化に向け、先行研究 [2] の実験手続きを参考に条件を統制したデータ収集法について検討した。また、前処理では、RVS と主観ストレス双方の個人差緩和および不均等なデータセットの均等化、さらに、データセットの課題別分割によって更なる精度向上の方法を検討した。そして、最後に、深層学習における特徴抽出能力の向上に向けた改善案について検討した。

第 4 章

評価と考察

4.1 緒言

本章では、3 クラス分類時における“Low-level-stress”の F 値の向上に対する改善手法について先行研究データ [2] と本実験データを対象とした評価結果を述べる．それにあたり、まず初めに、心理実験にて収集した参加者データの解析を行い、その後、改善手法を適用したストレスの推定精度の評価結果より今後の課題考察に触れる．

4.2 参加者データの解析

本研究では、心理実験の際に、主観ストレス評価に加え、各課題における正答率および 1 問当たりの反応速度のデータを収集した．そのため、これらのデータに対してそれぞれ解析を行い、実験参加者に対して効果的にストレスが誘発されていたかどうか評価した．

4.2.1 主観ストレス評価の解析

主観ストレス評価に関して、以下の 3 つの検定を有意水準 5%として実施した．

1. Test 直後の主観ストレス評価のみを対象とした場合における Math と Stroop それぞれの難易度間の平均値の差について
2. 1 セット目の Relax(主観評価の基準) 直後と各課題の easy 実施後の主観ストレス評価を対象とした場合の平均値の差について

4.2 参加者データの解析

3. 各セットの Relax 直後と Test 実施後の主観ストレス評価の差分を対象とした場合の Math と Stroop それぞれの難易度間の平均値の差について (Baseline 適用時の主観ストレス評価について)

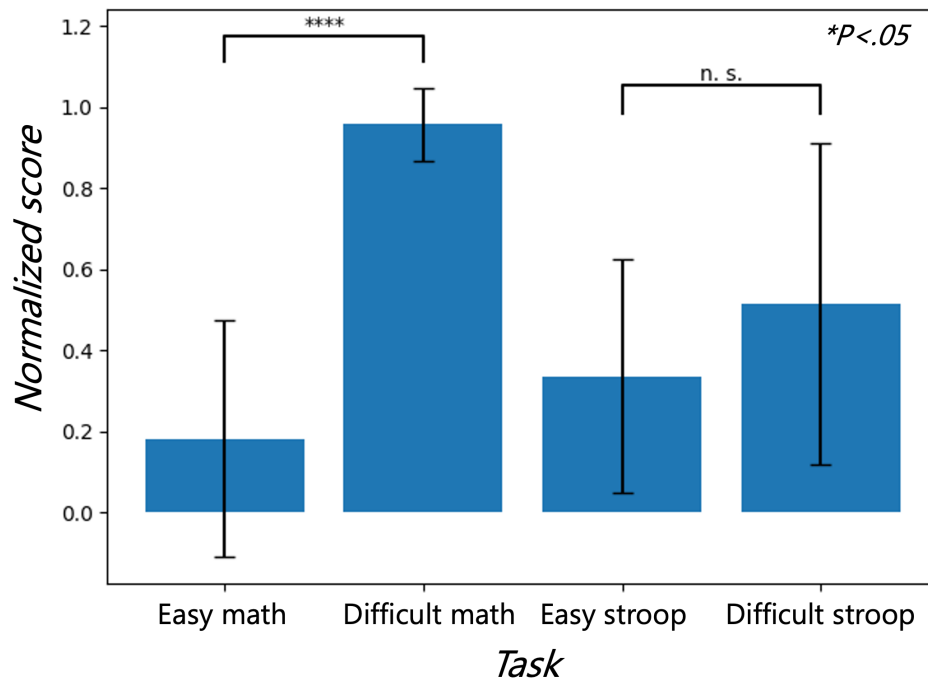


図 4.1: Test 直後の主観ストレス評価のみを対象とした場合における各課題難易度間の平均値の差についての検定結果

検定 1 の結果を図 4.1 に示す. Math は easy と difficult の 2 群間の平均値に有意差があった (*Easy math* – *Difficult math*: $t(10)=-8.78$, $p<.0001$) もの、Stroop では有意差が認められなかった (*Easy stroop* – *Difficult stroop*: $t(10)=-1.41$, $p=.186$). Test 直後の主観ストレス評価のみを検定の対象とした場合、個人別の基準が設けられていないため、客観的ストレスの評価とも言える. 先行研究 [2] では RVS に対するラベルを Test 直後のストレスを基に生成していたが、検定 1 の結果より、この方法では、各課題に対する主観ストレス評価を適切に実施できていないと考えられる.

4.2 参加者データの解析

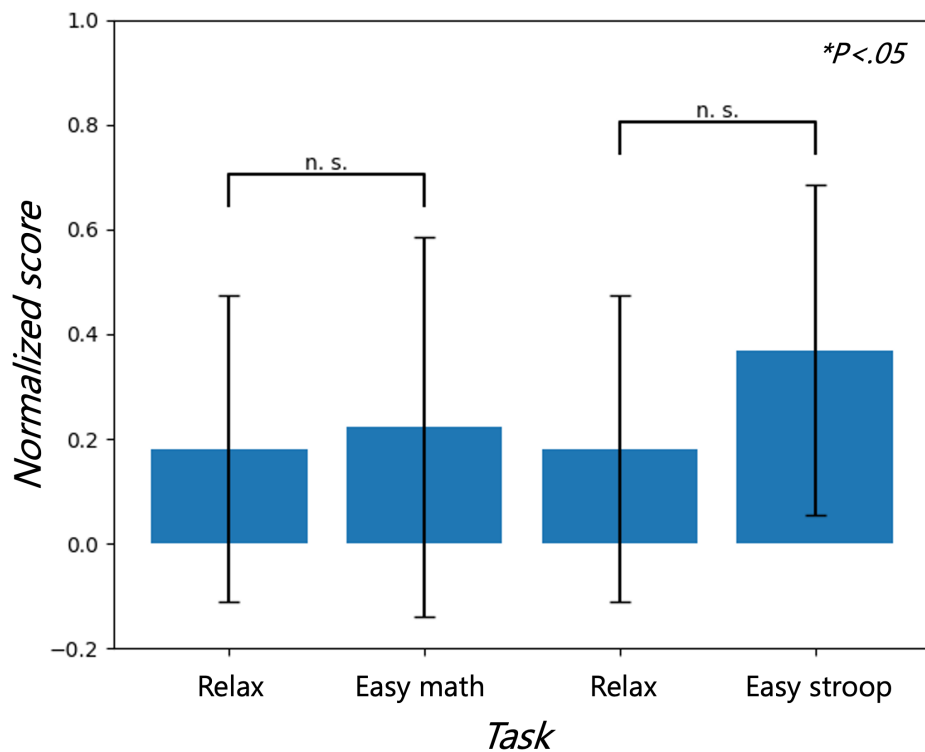


図 4.2: 1 セット目の Relax 直後と各課題の easy 実施後の主観ストレス評価についての検定結果

検定 2 の結果を図 4.2 に示す. Relax 直後と Easy math 実施後の主観ストレスおよび Relax 直後と Easy stroop 実施後の主観ストレスには有意差が認められなかった (*Relax* – *Easy math*: $t(10)=-0.74$, $p=.475$, *Relax* – *Easy stroop*: $t(10)=-0.80$, $p=.441$). この結果より, Math と Stroop の easy では, ストレスを誘発しておらず, Relax 直後の主観ストレス評価が個人の基準となり得ることが分かる.

検定 3 の結果を図 4.3 に示す. Math と Stroop は easy と difficult の 2 群間の平均値にそれぞれ有意差があった (*Easy math* – *Difficult math*: $t(10)=-5.65$, $p<.001$, *Easy stroop* – *Difficult stroop*: $t(10)=-3.73$, $p<.01$). Test 直前の Relax を主観ストレス評価の基準として Test 後との差分を算出することで, 各課題に対する純粋な主観ストレスを評価できていると考えられる. これを踏まえると, Math と Stroop 双方の課題にて difficult は easy に比べて効果的に実験参加者のストレスを誘発していることが分かる. さらに, 個人毎に主観ストレス評価の基準を設けたことで個人差が緩和され,

4.2 参加者データの解析

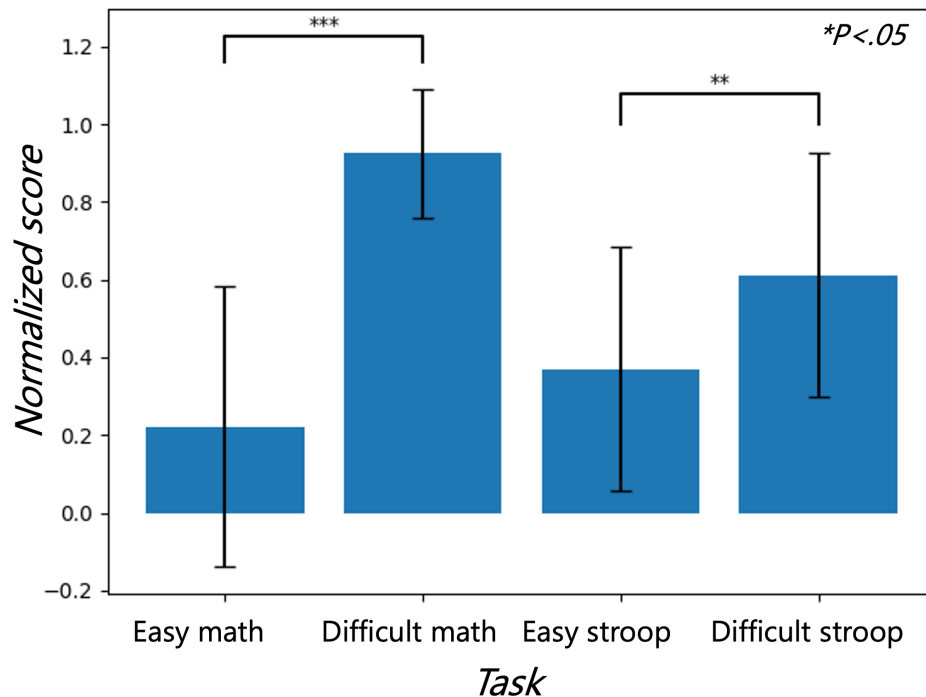


図 4.3: Relax 直後と Test 直後の主観ストレス評価の差分を対象とした場合における各課題難易度間の平均値の差についての検定結果

検定 1 の結果とは異なるように双方の課題難易度間に有意差が認められたと考えられる。

4.2.2 課題正答率および反応速度の解析

各課題難易度間における正答率および反応速度について、有意水準を 5% として検定した結果を図 4.4 と図 4.5 にそれぞれ示す。

前者に関して、Math と Stroop 双方にて easy に比べ、difficult では有意に正答率が低下している (*Easy math – Difficult math: $t(10)=5.60, p<.001$, Easy stroop – Difficult stroop: $t(10)=2.31, p<.05$*). しかし、図 4.4 から確認されるように、Difficult stroop はそれぞれの easy と同程度に正答率が高くおよそ 8 割強となっている。それに対して、Difficult math では、正答率の平均値が 6 割を下回る。後者の課題難易度別反応速度の検定結果より、Math では easy と difficult の 2 群間に有意差が見

4.3 ストレスの推定精度

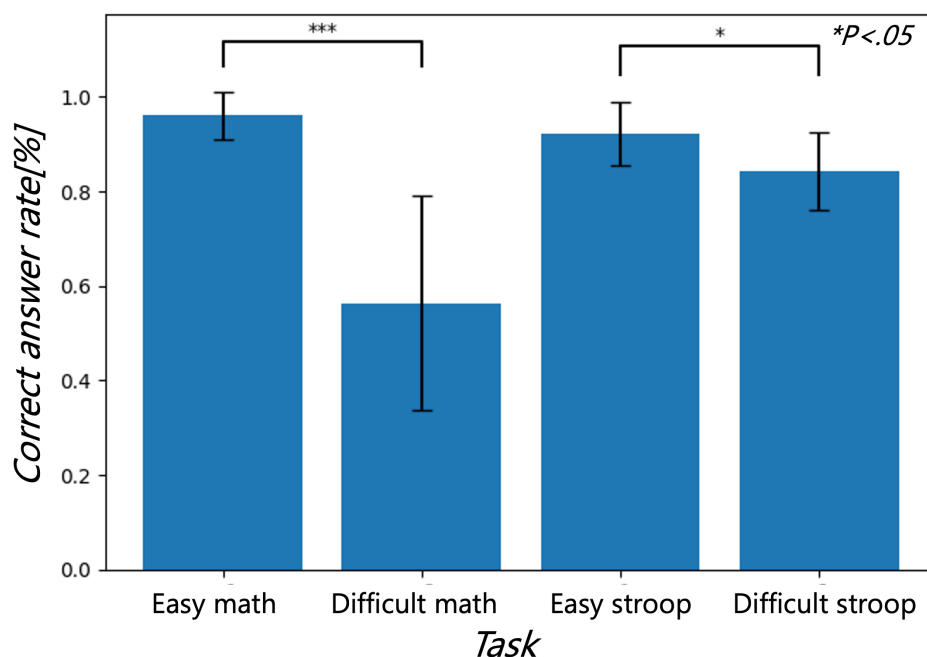


図 4.4: 課題難易度別正答率の検定結果

られたものの、Stroop では有意差は見られなかった (*Easy math* – *Difficult math*: $t(10)=-8.71$, $p<.0001$, *Easy stroop* – *Difficult stroop*: $t(10)=1.86$, $p=.093$).

以上より、課題の難易度に比例して正答率は低下し、特に Difficult math のように認知負荷が大きい課題では反応速度が低下することが確認された。これらの結果から主観ストレス強度と課題正答率および反応速度には関連性が見られる。

4.3 ストレスの推定精度

3 クラス分類時における “Low-level-stress” の F 値の向上に対する改善手法について追検証時に使用した同様の深層学習モデルによって評価した。データセットには先行研究データ [2] と本実験により収集したデータを使用した。先行研究データ [2] に関して、既に RVS と主観ストレスによるラベルが対応付けられた状態で配布されており、前処理の改善手法を適用できなかったため、**Augmented** と **Balanced** のみ評価を行った。評

4.3 ストレスの推定精度

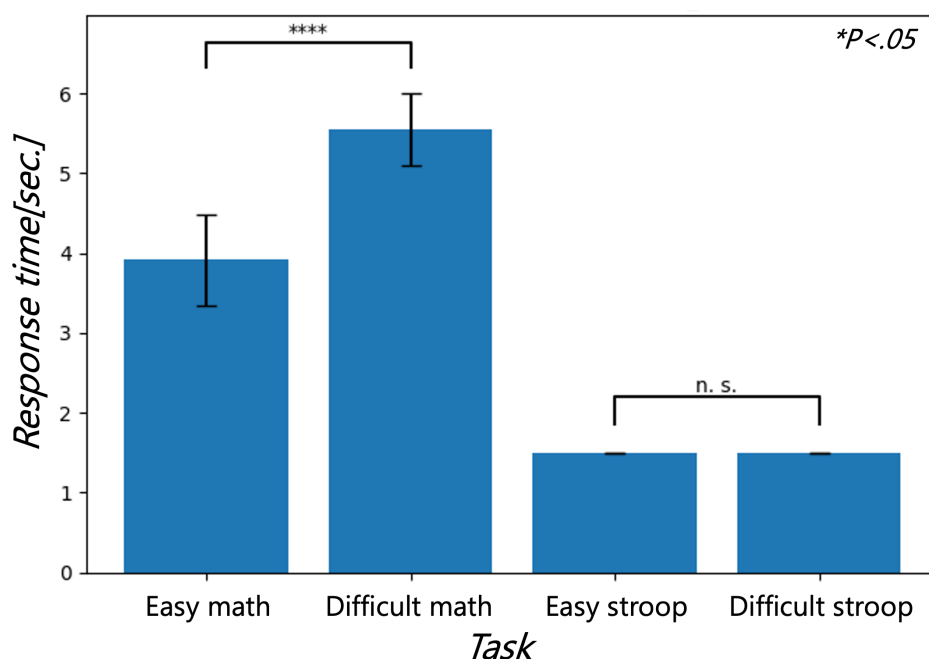


図 4.5: 課題難易度別反応速度の検定結果

価方針として、まず初めに、それぞれのデータセットを対象に各手法単独の有効性について検証し、その後、有効性の高い組み合わせについて評価を行った。また、評価方法は先行研究 [2] と同様に Loso 交差検証法とした。

4.3.1 各手法単独の有効性についての検証

〈 先行研究データ [2] を対象にした評価 〉

- **Augmented** の結果

2 層 2D-CNN による結果を表 4.1 に示す。適用した場合、そうでない場合に比べ、2 クラス分類と 3 クラス分類時の精度が共に 6%程低下する結果となった。そこで、水平方向と垂直方向の反転をそれぞれ単独で実施した場合の評価も行った。結果を表 4.2 に示すように、水平方向のみ適用した場合は、2 クラス分類と 3 クラス分類時の精度は Augmented を適用しない場合より 1~3%程低下した。また、垂直方向の

4.3 ストレスの推定精度

みの反転では、さらに精度が低下し、特に 3 クラス分類時の精度は水平方向のみの反転に比べて 8%程低下する結果となった。これらの結果を踏まえると、特に垂直方向の反転を適用した場合、RVS の特徴を上手く学習できないため、精度向上への有効性が低いと考えられる。そのため、本実験データには Augmented を適用せず、その他の 5 つの手法について評価する。

表 4.1: Augmented の適用の有無の比較

	データ拡張無	データ拡張有
2 クラス分類時	70.81	65.19
3 クラス分類時	55.79	48.83

表 4.2: 水平方向と垂直方向の比較

	水平方向	垂直方向
2 クラス分類時	69.61	65.27
3 クラス分類時	52.67	44.77

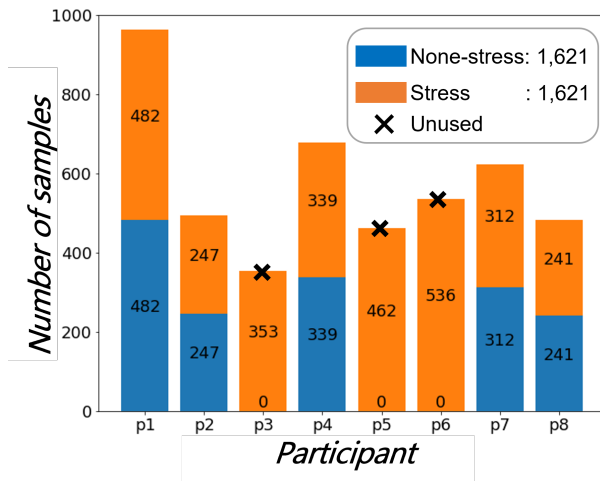
● **Balanced** の結果

初めに、図 4.6 に Oversampling と Down-sampling の 2 通りで Balanced を適用した際の 2 クラス分類時と 3 クラス分類時のラベルの分布を示す。3.4 節にて先述したように評価したい全てのストレスクラスのラベルを保有していない参加者のデータは学習とテストに使用しない。2 クラス分類時の場合、p3, p5, p6 は “Stress” のラベルしか保有していないため、残りの p1, p2, p4, p7, p8 の 5 人分のデータのみ使用する。また、3 クラス分類時の場合についても同様に、p2 と p7 以外は 3 つ全てのストレスクラスのラベルを保有していないため、学習とテストには使用しない。

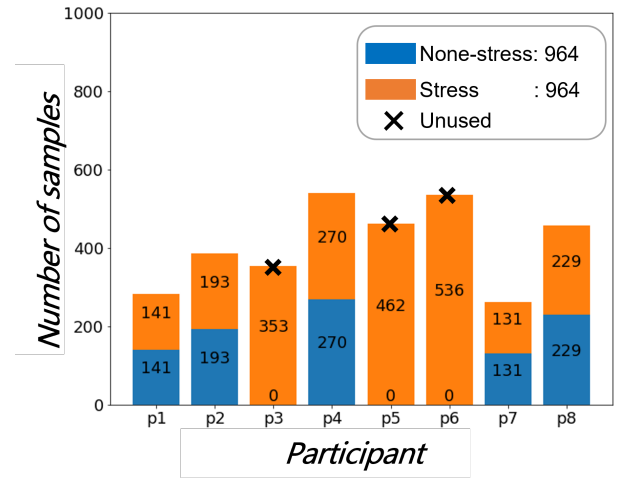
－ Oversampling 適用時

2 クラス分類時と 3 クラス分類時の精度と各ストレスクラスの F 値を表 4.3 と表 4.4 に示す。2 クラス分類時の精度はいずれも低下する結果となった。一方で、

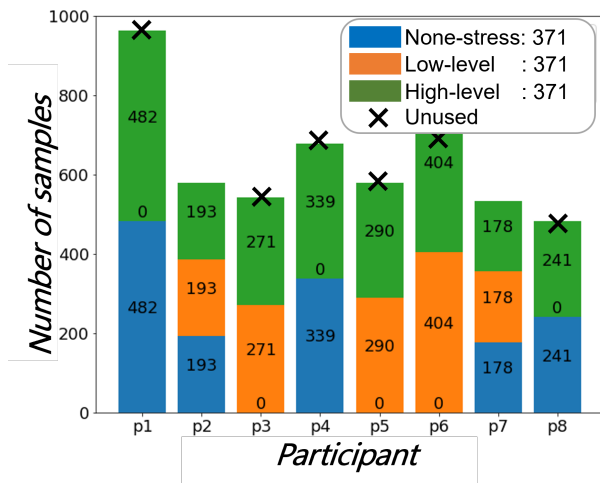
4.3 ストレスの推定精度



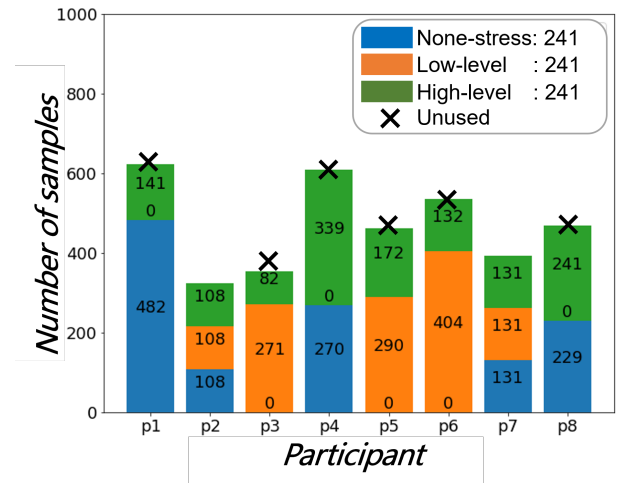
(a) 2 クラス分類時における Oversampling 適用後



(b) 2 クラス分類時における Down-sampling 適用後



(c) 3 クラス分類時における Oversampling 適用後



(d) 3 クラス分類時における Down-sampling 適用後

図 4.6: 先行研究データ [2] に Balanced を適用したラベルの分布

“None-stress” の F 値は全体的に向上し、最大で EfficientNet-b0[17] の 74.0% となった。3 クラス分類時の精度は半数のモデルにて向上したものの、“Low-level-stress” の F 値は全体的に低下し、特に MobileNet-v3[16] と EfficientNet-b0[17] による転移学習では 1.3%~1.6% と極端に低い値となった。

2 クラス分類時の結果より、不均等データでは、“None-stress” の F 値が低く、“Stress” が高い値であったのに対し、均等データでは双方の F 値の差が縮小したことを踏まえると、いずれの学習モデルも RVS の特徴よりラベルの頻度を学

4.3 ストレスの推定精度

習している可能性がある。

表 4.3: 先行研究データ [2] を対象とした Over 適用時の 2 クラス分類の精度と F 値

	不均等データ (8 subjects)			均等データ (5 of 8 subjects)		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
2 層 2D-CNN[2]	70.81	44.89	80.15	68.26	64.31	71.42
3 層 2D-CNN	68.01	20.16	80.00	62.71	40.53	72.84
5 層 2D-CNN	66.79	40.29	77.00	60.06	43.87	69.00
ResNet18[15]	71.95	40.65	81.64	64.24	51.99	71.48
MobileNet-v3[16]	82.72	71.36	87.63	74.65	69.85	78.13
EfficientNet-b0[17]	80.41	67.48	85.98	77.14	73.97	79.63

表 4.4: 先行研究データ [2] を対象とした Over 適用時の 3 クラス分類の精度と F 値

	不均等データ (8 subjects)				均等データ (2 of 8 subjects)			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	55.79	57.39	43.89	63.96	43.94	52.85	13.22	66.10
3 層 2D-CNN	45.75	34.22	45.57	51.29	51.21	64.58	44.83	47.37
5 層 2D-CNN	55.49	58.73	37.59	65.89	31.18	58.59	22.91	14.48
ResNet18[15]	41.34	49.49	5.37	60.86	36.57	35.71	7.49	61.56
MobileNet-v3[16]	43.80	66.67	9.00	50.17	55.26	67.56	1.59	74.08
EfficientNet-b0[17]	44.99	63.23	7.52	54.57	52.56	63.69	1.25	69.42

4.3 ストレスの推定精度

– Down-sampling 適用時

2 クラス分類時と 3 クラス分類時の精度と各ストレスクラスの F 値を以下の表 4.5 と表 4.6 に示す.

表 4.5: 先行研究データ [2] を対象とした Down 適用時の 2 クラス分類の精度と F 値

	不均等データ (8 subjects)			均等データ (5 of 8 subjects)		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
2 層 2D-CNN[2]	70.81	44.89	80.15	52.07	20.07	65.78
3 層 2D-CNN	68.01	20.16	80.00	50.00	0.00	66.67
5 層 2D-CNN	66.79	40.29	77.00	50.00	0.00	66.67
ResNet18[15]	71.95	40.65	81.64	50.26	0.00	66.78
MobileNet-v3[16]	82.72	71.36	87.63	50.00	0.00	66.67
EfficientNet-b0[17]	80.41	67.48	85.98	50.00	0.00	66.67

表 4.6: 先行研究データ [2] を対象とした Down 適用時の 3 クラス分類の精度と F 値

	不均等データ (8 subjects)				均等データ (2 of 8 subjects)			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	55.79	57.39	43.89	63.96	59.14	69.11	51.67	56.08
3 層 2D-CNN	45.75	34.22	45.57	51.29	58.86	55.98	45.39	72.16
5 層 2D-CNN	55.49	58.73	37.59	65.89	66.95	74.21	68.68	55.87
ResNet18[15]	41.34	49.49	5.37	60.86	60.25	75.83	26.50	66.19
MobileNet-v3[16]	43.80	66.67	9.00	50.17	35.15	16.67	0.00	58.60
EfficientNet-b0[17]	44.99	63.23	7.52	54.57	37.52	0.60	1.92	56.21

2 クラス分類時の精度はいずれも低下し, 2 層 2D-CNN[2] を除く, いずれのモ

4.3 ストレスの推定精度

デルにおいてチャンスレベルとなり分類不可であった。また、F 値に関して、いずれも低下し、2 層 2D-CNN[2] 以外では“None-stress”が 0%となり、“Stress”の特徴しか学習できていない結果となった。その一方で、3 クラス分類時の精度は 2D-CNN と ResNet[15] にて向上し、特に 5 層 2D-CNN では、最も高い 67.0%となった。また、F 値に関して“Low-level-stress”では半数のモデルにおいて向上したが、MobileNet-v3[16] と EfficientNet-b0[17] による転移学習では 0.0~1.9%とさらに低下する結果となった。

上記の Oversampling と Down-sampling の結果より、不均等なデータセットの均等化は精度向上に対して必ずしも有効な手法とは言えない。特に、Down-sampling 適用時における 2 クラス分類時の精度は、2 層 2D-CNN[2] を除く全てのモデルにおいてチャンスレベルとなり、RVS に基づくストレス推定が不可能であった。一方、3 クラス分類時は、一部のモデルにおいて精度と“Low-level-stress”の F 値に改善が見られた。しかし、2.4 節にて述べたように、RVS には“Low-level-stress”固有の特徴が含まれていないと仮定した場合、この結果には整合性がない。また、Down-sampling 同様に、Oversampling 適用時には、精度の改善が見られなかったが、各ストレスクラスの F 値に関して、先述した RVS に含まれる特徴の仮説と合致することから、適切に評価できている可能性が高い。さらに、不均等なデータセットでは、ラベルの偏りを踏まえると均等化したデータセットとの比較では、F 値が評価指標になると考えられる。

〈 本実験データを対象にした評価 〉

- **Baseline**

2 層 2D-CNN[2] による結果を表 4.7 に示す。Baseline を適用しない場合、2 クラス分類ではチャンスレベルを下回り、分類不可であったが、適用する場合は 2 クラス分類と 3 クラス分類双方において分類可能となった。これらの結果は、図 4.1 と図 4.3 に示した Math と Stroop 双方の 2 つの難易度間に有意差があるか否かに関連

4.3 ストレスの推定精度

していると考えられる。実際、主観ストレスに対する基準を設けた場合、個人差が緩和されていたことで双方の課題において有意差が認められ、分類精度に関しても Baseline 無しよりも 10%程高い値となった。

この結果より、Baseline は分類精度向上への有効性が高いと考えられるため、本実験により収集したデータには常時適用した主観ストレスのラベルを用いることとする。

表 4.7: Baseline の適用の有無の比較

	Baseline 無	Baseline 有
2 クラス分類時	46.75	57.10
3 クラス分類時	38.22	48.37

- Cut

2 クラス分類時と 3 クラス分類時における結果を以下の表 4.8 と表 4.9 に示す。2 クラス分類時の精度は半数のモデルにて向上し、2 層 2D-CNN[2] の 61.1%が最も高い値となった。また、F 値に関して、“None-stress” は最大で 70.4%となったものの、2 層 2D-CNN[2] と ResNet18[15] 以外では低下し、“Stress” の値についてはいずれも低下する結果となった。3 クラス分類時の精度は 3 層と 5 層の 2D-CNN を除く、半数以上のモデルにて向上し、ResNet18[15] の 52.0%が最も高い値となった。また、“Low-level-stress” の F 値についても半数のモデルにて向上する結果となった。

この結果より、課題開始冒頭 30 秒間は緊張等により呼吸が安定しておらず、不安的期間のデータを削除することによって個人毎に一貫した RVS が生成されている。つまり、RVS における個人差が緩和されていると考えられるため、分類精度向上への有効性が期待される。

4.3 ストレスの推定精度

表 4.8: Cut 適用時の 2 クラス分類の精度と F 値

	Cut 無			Cut 有		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
2 層 2D-CNN[2]	57.10	66.26	41.14	61.13	70.43	43.31
3 層 2D-CNN	57.14	65.73	42.78	47.39	57.05	32.13
5 層 2D-CNN	56.02	65.24	40.15	52.99	63.23	34.85
ResNet18[15]	56.91	66.30	40.27	57.24	66.90	39.60
MobileNet-v3[16]	53.32	66.36	35.71	55.42	65.22	37.92
EfficientNet-b0[17]	59.61	67.77	46.09	56.62	66.78	37.50

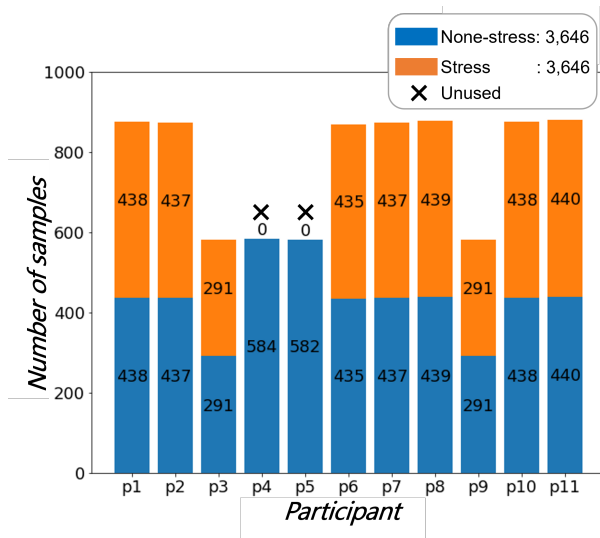
表 4.9: Cut 適用時の 3 クラス分類の精度と F 値

	Cut 無				Cut 有			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	48.37	40.10	46.16	57.16	49.70	42.41	47.80	56.56
3 層 2D-CNN	50.22	40.18	48.87	60.37	32.47	39.64	34.83	24.63
5 層 2D-CNN	45.42	27.81	45.99	57.58	42.82	31.40	41.89	52.69
ResNet18[15]	42.43	33.58	43.77	48.26	51.95	43.33	55.95	54.62
MobileNet-v3[16]	36.80	32.58	35.45	41.45	42.92	39.18	36.75	50.26
EfficientNet-b0[17]	37.35	39.39	26.38	44.82	38.53	37.51	24.49	48.95

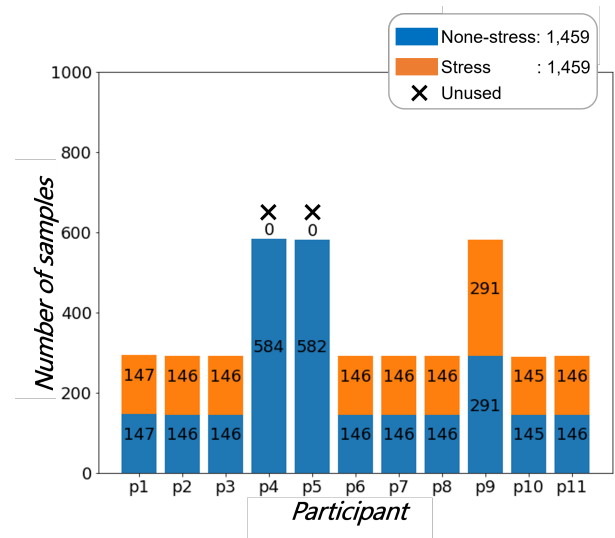
4.3 ストレスの推定精度

• Balanced

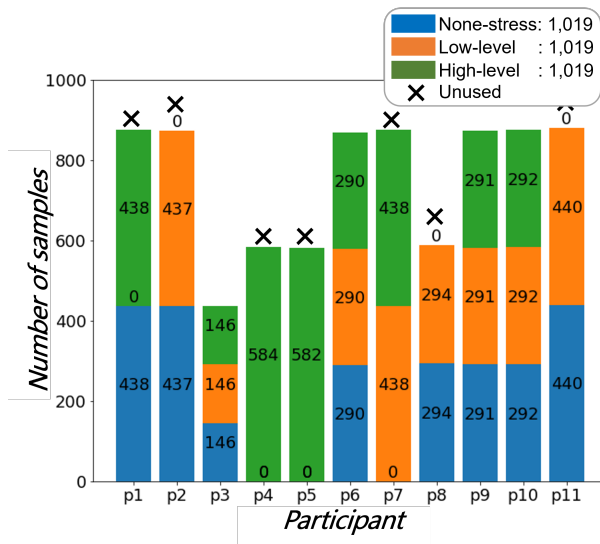
初めに, 図 4.7 に Oversampling と Down-sampling の 2 通りで Balanced を適用した際の 2 クラス分類時と 3 クラス分類時のラベルの分布を示す. 2 クラス分類時には, p4 と p5 を除く 9 人分のデータを, 3 クラス分類時には, p3 と p6, p9, p10 の 4 人分のデータを学習とテストに使用した. 2 クラス分類時および 3 クラス分類時における各ストレスクラスのラベル数は図 4.7 の a~d に示すように, それぞれ異なる.



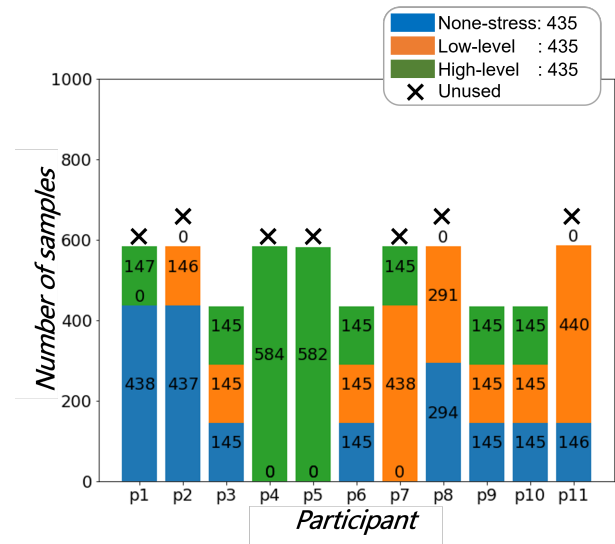
(a) 2 クラス分類時における Oversampling 適用後



(b) 2 クラス分類時における Down-sampling 適用後



(c) 3 クラス分類時における Oversampling 適用後



(d) 3 クラス分類時における Down-sampling 適用後

図 4.7: 本実験データに Balanced を適用したラベルの分布

4.3 ストレスの推定精度

– Oversampling 適用時

2 クラス分類時と 3 クラス分類時の精度と各ストレスクラスの F 値を表 4.10 と表 4.11 に示す.

表 4.10: 本実験データを対象とした Over 適用時の 2 クラス分類の精度と F 値

	不均等データ (11 subjects)			均等データ (9 of 11 subjects)		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
2 層 2D-CNN[2]	57.10	66.26	41.14	56.21	60.35	51.11
3 層 2D-CNN	57.14	65.73	42.78	54.21	56.45	51.73
5 層 2D-CNN	56.02	65.24	40.15	54.42	61.04	45.08
ResNet18[15]	56.91	66.30	40.27	45.90	52.94	36.38
MobileNet-v3[16]	53.32	66.36	35.71	47.42	52.68	40.85
EfficientNet-b0[17]	59.61	67.77	46.09	54.14	58.25	49.13

表 4.11: 本実験データを対象とした Over 適用時の 3 クラス分類の精度と F 値

	不均等データ (11 subjects)				均等データ (4 of 11 subjects)			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	48.37	40.10	46.19	57.16	45.83	38.35	51.98	46.46
3 層 2D-CNN	50.22	40.18	48.87	60.37	49.26	39.22	44.43	63.79
5 層 2D-CNN	45.42	27.81	45.99	57.58	42.53	36.91	41.00	48.73
ResNet18[15]	42.43	33.58	43.77	48.26	38.11	15.61	46.66	43.86
MobileNet-v3[16]	36.80	32.58	35.45	41.50	45.57	41.65	46.78	47.65
EfficientNet-b0[17]	37.35	39.39	26.38	44.82	35.95	26.01	34.52	44.65

2 クラス分類時の精度はいずれも低下し, “None-stress” の F 値に関しても低下

4.3 ストレスの推定精度

する結果となった。3 クラス分類時の精度は MobileNet-v3[16] でのみ向上したものの，“Low-level-stress” の F 値は 3 層と 5 層の 2D-CNN を除くモデルにおいて向上する結果となった。

– Down-sampling 適用時

2 クラス分類時と 3 クラス分類時の精度と各ストレスクラスの F 値を表 4.12 と表 4.13 に示す。2 クラス分類時の精度は ResNet18[15] と EfficientNet-b0[17] にて向上した。“None-stress” の F 値は 3 層 2D-CNN のみ向上したが，その他いずれにおいても低下し，同時に “Stress” についても低下する結果となった。3 クラス分類時の精度は 3 層と 5 層の 2D-CNN を除くモデルにて向上した。また，“Low-level-stress” の F 値に関して，半数のモデルにおいて改善が見られたが，3 層 2D-CNN では 0.0%となった。

表 4.12: 本実験データを対象とした Down 適用時の 2 クラス分類の精度と F 値

	不均等データ (11 subjects)			均等データ (9 of 11 subjects)		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
2 層 2D-CNN[2]	57.10	66.26	41.14	56.48	62.49	48.25
3 層 2D-CNN	57.14	65.73	42.78	51.95	67.55	7.52
5 層 2D-CNN	56.02	65.24	40.15	53.22	64.26	32.33
ResNet18[15]	56.91	66.30	40.27	59.60	63.05	55.42
MobileNet-v3[16]	53.32	66.36	35.71	47.74	64.61	0.01
EfficientNet-b0[17]	59.61	67.77	46.09	59.90	61.08	58.66

上記の Oversampling と Down-sampling の結果より，後者では，2 クラス分類時の “Stress” の F 値が極端に低下し，また，3 クラス分類時には “Low-level-stress” の F 値が 0.0%になる場合があった。それに対して，前者の Oversampling では，各ストレスクラスの F 値が極端に低下する場合がないことから，分類精度を適切に評価で

4.3 ストレスの推定精度

表 4.13: 本実験データを対象とした Down 適用時の 3 クラス分類の精度と F 値

	不均等データ (11 subjects)				均等データ (4 of 11 subjects)			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	48.37	40.10	46.19	57.16	52.59	49.76	55.13	52.77
3 層 2D-CNN	50.22	40.18	48.87	60.37	33.33	28.57	0.00	46.15
5 層 2D-CNN	45.42	27.81	45.99	57.58	39.71	21.41	40.03	52.58
ResNet18[15]	42.43	33.58	43.77	48.26	53.91	54.40	51.38	56.06
MobileNet-v3[16]	36.80	32.58	35.45	41.50	37.87	38.32	44.53	27.47
EfficientNet-b0[17]	37.35	39.39	26.38	44.82	49.71	54.27	29.07	58.14

きている可能性がある。そのため、Oversampling による Balanced を用いた評価が必須となる。

- **Separated**

表 4.14 に 3 クラス分類時の結果を示す。Math の精度は 39.5%~54.0%で、Stroop は 37.9%~52.7%と類似する結果となった。F 値に関して、Math の “Low-level-stress” は 0.0%~7.4%と極端に低い値であった一方、“High-level-stress” は 3 つのストレスクラスの中で最も高い値となった。また、Stroop では対極の結果となり、“Low-level-stress” が最も高い値となった。以上の結果より、Math 実行時の RVS には “High-level-stress” の特徴が多く含まれ、Stroop 実行時には “Low-level-stress” の特徴が多く含まれている可能性が高い。

しかし、図 4.8 に示す Separated 適用時のラベルの分布から確認されるよう、実際、Math のデータセットは “High-level-stress” のラベルが半数以上を占めており、“Low-level-stress” が極端に少ない。また、Stroop では、“None-stress” と “High-level-stress” に比べて “Low-level-stress” が極端に多く、半数以上を占めている。このようなラベルの偏りによって Math と Stroop では F 値の値が対極の結果となっ

4.3 ストレスの推定精度

表 4.14: Separated 適用時の 3 クラス分類の精度と F 値

	Math				Stroop			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	43.39	22.08	0.00	58.73	40.21	27.28	55.72	0.00
3 層 2D-CNN	49.49	24.20	0.00	63.85	41.96	17.26	58.35	0.00
5 層 2D-CNN	39.49	17.41	0.00	55.45	37.84	19.83	54.04	7.11
ResNet18[15]	49.40	29.78	0.65	63.77	52.71	29.76	67.29	5.15
MobileNet-v3[16]	47.86	32.30	7.43	60.57	38.97	11.94	51.34	26.78
EfficientNet-b0[17]	53.97	36.86	5.00	65.81	48.63	8.84	63.27	23.66

ていたと考えられる。そのため、この Separated 適用時には各ストレスクラスの適切な分類精度を評価できない。

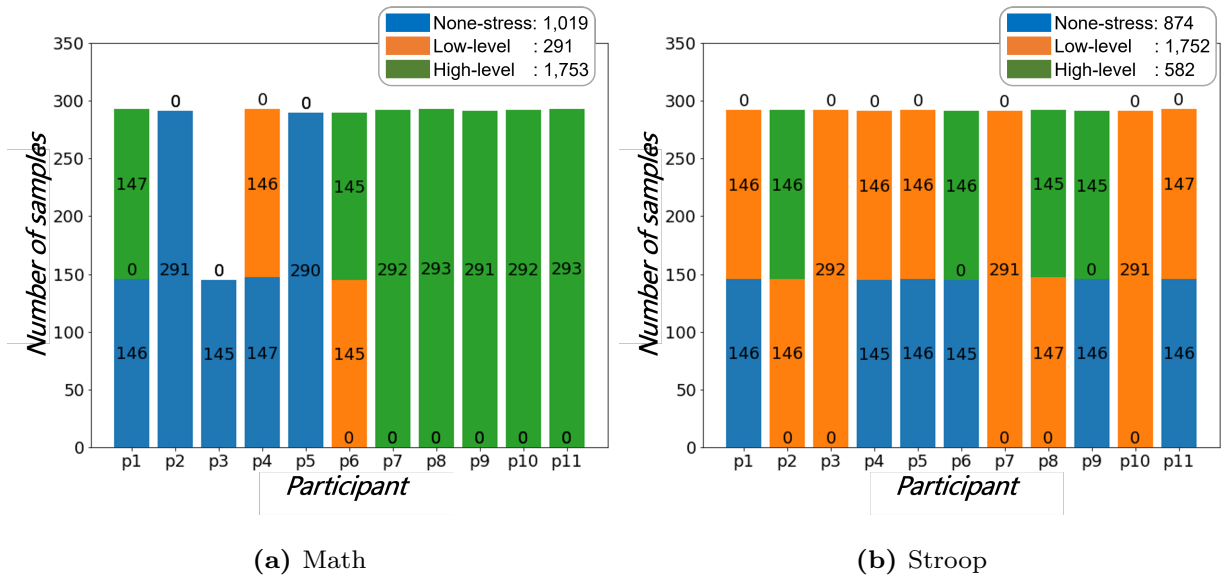


図 4.8: Separated 適用時の 3 クラス分類時のラベル分布

- Transfer learning

表 4.15 と表 4.16 に転移学習による結果を示す。

4.3 ストレスの推定精度

表 4.15: Transfer learning 適用時の 2 クラス分類の精度と F 値

	Xfer 無			Xfer 有		
	精度	F 値		精度	F 値	
		None	Stress		None	Stress
ResNet18[15]	56.91	66.30	40.27	59.64	69.03	42.10
MobileNet-v3[16]	53.32	66.36	35.71	57.33	66.50	41.24
EfficientNet-b0[17]	59.61	67.77	46.09	61.25	70.29	44.32

表 4.16: Transfer learning 適用時の 3 クラス分類の精度と F 値

	Xfer 無				Xfer 有			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
ResNet18[15]	42.43	33.58	43.77	48.26	45.29	39.35	46.37	49.57
MobileNet-v3[16]	36.80	32.58	35.45	41.45	40.95	36.61	36.45	48.27
EfficientNet-b0[17]	37.35	39.39	26.38	44.82	30.06	40.01	12.92	35.13

2 クラス分類時はいずれも精度が向上し、最大で EfficientNet-b0[17] は 61.3% となった。また、“None-stress” の F 値についても向上する結果となった。3 クラス分類時では、ResNet-18[15] と MobileNet-v3[16] にて精度の向上が見られたものの、EfficientNet-b0[17] ではチャンスレベルを下回り、分類ができていない。“Low-level-stress” の F 値についても同様に、ResNet-18[15] と MobileNet-v3[16] では改善したものの、EfficientNet-b0[17] では 12.9% と低下した。

EfficientNet-b0[17] のみ、3 クラス分類時の精度が改善されていない原因として、ResNet-18[15] と MobileNet-v3[16] よりもさらに多層となっているため、先行研究データ [2] のデータ不足に起因してモデルのファインチューニングが十分に行えていなかった可能性がある。その一方で、他 2 つのモデルでは、精度と F 値が向上し

4.3 ストレスの推定精度

たため、有効性が高いと考えられる。

以上の各手法単独での分類精度および F 値向上への有効性について ○, △, × の 3 段階で評価した結果を以下の表 4.17 にまとめる。まず, Baseline と Cut, Transfer learning についてはいずれも 2 クラス分類および 3 クラス分類時においてそれぞれ適用しない場合に比較して精度と F 値に改善が見られた。また, Balanced の Oversampling では, 精度と F 値が向上する場合もあり, 加えて, 各ストレスクラスの分類精度を適切に評価する方法として必須であると考えられる。その一方で, Balanced の Down-sampling では, 学習データの減少等によって, 2 クラス分類時では, “Stress” の F 値と 3 クラス分類時では, “Low-level-stress” の F 値が極端に低下する場合があります, 分類精度を適切に評価することができていないと考えられる。さらに, Separated と Augmented では, それぞれ, 極端なラベルの偏りに起因して分類精度を適切に評価できず, 分類精度が 10%程低下する結果となった。これらの結果を基に次節では, 有効性の高いと考えられる Baseline と Cut, Balanced(Oversampling), Transfer learning 4 つの複数手法を組み合わせるについて検証する。

表 4.17: 各手法単独の有効性についての結果

提案手法		有効性
Baseline		○
Cut		○
Balanced	Oversampling	△
	Down-sampling	×
Separated		×
Augmented		×
Transfer learning		○

4.3 ストレスの推定精度

4.3.2 複数の手法を組み合わせた有効性についての検証

前節では、RVSに基づくストレス推定の精度向上を目的に条件を統制した心理実験によるデータ収集と個人差を緩和させる前処理等、そして、深層学習モデルの特徴抽出能力の向上といった大きく3つの改善手法についてそれぞれ単独の有効性を検証した。結果として、Baseline と Cut, Balanced(Oversampling), Transfer learning では精度および3クラス分類時における“Low-level-stress”のF値向上への有効性が高いと考えられるため、この4つ全てを組み合わせた手法について検証した。また、この際、比較対象として、先行研究データ [2] に Balanced(Oversampling) のみ適用した手法を用いた。結果を以下の表 4.18 に示す。

表 4.18: 本実験データに Base+Cut+Bal.+Xfer を組み合わせた3クラス分類時の精度とF値の比較

	先行研究データ [2]				本実験データ			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
ResNet18[15]	36.57	35.71	7.49	61.56	35.14	17.80	13.50	49.31
MobileNet-v3[16]	55.26	67.56	1.59	74.08	27.53	20.65	7.68	39.80
EfficientNet-b0[17]	52.56	63.69	1.25	69.42	32.88	17.98	2.97	49.34

本実験データに関して、“Low-level-stress”のF値はいずれも向上したが、先行研究データ [2] に Balanced(Oversampling) のみ適用した場合と比較してその差は僅かであった。また、精度に関しては、先行研究データ [2] よりも低下し、期待する結果は得られなかった。そこで、本研究では、Baseline と Cut, Balanced(Oversampling), Transfer learning を3つまたは、2つと組みわせて同様に検証したところ、本実験データに Baseline と Balanced(Oversampling) のみ適用した場合が最も良い結果となった。以下の表 4.19 に3クラス分類時における精度と各ストレスクラスのF値を示す。

先程同様に、先行研究データ [2] に Balanced のみ適用した場合と比較したところ、

4.4 考察

表 4.19: 本実験データに Base+Bal. を組み合わせた 3 クラス分類時の精度と F 値の比較

	先行研究データ [2](Bal.)				本実験データ (Base+Bal.)			
	精度	F 値			精度	F 値		
		None	Low	High		None	Low	High
2 層 2D-CNN[2]	43.94	52.85	13.22	66.10	45.83	38.35	51.98	46.46
3 層 2D-CNN	51.21	64.58	44.83	47.37	49.26	39.22	44.43	63.79
5 層 2D-CNN	31.18	58.59	22.91	14.48	42.53	36.91	41.00	48.73
ResNet18[15]	36.57	35.71	7.49	61.56	38.11	15.61	46.66	43.86
MobileNet-v3[16]	55.26	67.56	1.59	74.08	45.57	41.65	46.78	47.65
EfficientNet-b0[17]	52.56	63.69	1.25	69.42	35.95	26.01	34.52	44.65

精度に関して、本実験データでは、半数のモデルにて向上した。また、F 値に関して、“Low-level-stress” では、3 層 2D-CNN を除くモデルにおいて大幅に向上する結果となった。しかし、依然として、先行研究データ [2] の方が精度が高く、考えられる原因として、先行研究データ [2] に Balanced を適用した場合、8 人中 2 人分のデータしか学習とテストに使用しておらず、偶然にも個人差による影響が小さかった可能性がある。

4.4 考察

本研究では、RVS に基づくストレス推定における 3 クラス分類時の精度向上にあたり、改善すべき “Low-level-stress” の F 値に着目して提案手法の検討および評価を行った。結果として、本実験により収集したデータに対して主観ストレス強度のベースライン化による正規化 (Baseline) と不均等なデータセットの均等化 (Balanced) をすることで先行研究データ [2] に比較して “Low-level-stress” の F 値が大幅に向上し、特に MobileNet-v3[16] では、最大で 45%向上した。一方、精度に関しては期待する結果が得られなかったため、今後の課題として、本実験データに対して Baseline と Balanced を適

4.4 考察

用した際の“Low-level-stress”のF値を維持したまま精度の向上が必要である．本研究では，心理実験を実施した際，ストレスの他，疲労度・快不快・覚醒眠気についてもデータを収集しており，RVSに基づいてそれぞれの感情の推定精度を検証した結果，各感情に対する課題の難易度間の平均値に有意差がある場合のみ推定可能であった．実際に，主観ストレスと同様に有意水準を5%として，MathとStroopの各課題の難易度間(easyとdifficult)について対応のあるt検定によって解析した結果を以下の表4.20を示す．

表 4.20: 感情に対する課題難易度間の検定結果 (有意水準 5%)

	Math (easy – diff)	Stroop (easy – diff)
ストレス	$t(10) = -5.65$	$t(10) = -3.73$
疲労度	$t(10) = -2.85$	$t(10) = -1.59$
快不快	$t(10) = -3.43$	$t(10) = -1.15$
覚醒眠気	$t(10) = 4.68$	$t(10) = 2.43$

(太字： p 値 $< .05$)

Mathに関して，ストレスとその他いずれの感情においても有意差が見られ，この結果より，2段階のMathは，効果的に様々な感情を誘発させることが分かる．Stroopでは，ストレスと覚醒眠気にて有意差が見られたものの，疲労度と快不快については有意差が認められず，Mathに比べて効果的に感情を誘発できていないと考えられる．これらの結果を踏まえ，ImageNetにて事前学習済みの3つの既存モデル(ResNet18[15], MobileNet-v3[16], EfficientNet-b0[17])に基づく転移学習によって，分類精度を検証した結果を表4.21に示す．表4.21から確認されるよう，ストレスと覚醒眠気のようにMathとStroop双方で2つの課題難易度間(easyとdifficult)に有意差がある場合のみ，いずれのモデルにおいても2クラス分類と3クラス分類を正しく行うことができた．その一方で，疲労度と快不快のように，MathとStroop双方で有意差が認められない感情に関しては，一部のモデルにて2クラス分類時と3クラス分類時の精度がチャンスレベルを超えたものの，大部分では分類不可となった．これらの結果から，RVSに基づくストレ

4.4 考察

ス推定の精度は心理実験にて収集するデータの質に依存する可能性が高いと考えられる。

表 4.21: 3つの既存モデルによる感情の分類精度 (%)

	2 クラス分類時	3 クラス分類時
ストレス	71.95~82.72	41.34~44.99
疲労度	48.37~ 51.89	30.55~ 35.57
快不快	42.51~45.11	24.19~32.42
覚醒眠気	52.77~61.36	38.80~47.49

本研究では、RVS に基づくストレス推定の精度向上へ向け、心理実験によるデータ収集と前処理、そして、深層学習によるストレス推定、それぞれに対して改善策を講じた。先述した各感情と課題難易度間の解析結果および分類精度の結果より、RVS に基づくストレスあるいは、感情の推定精度は心理実験にて収集するデータの質に依存すると考えられる。そのため、今後は、更なる条件の統制によって心理実験を実施し、データを収集することが最重要課題となる。具体的に、本実験の結果より、Math は効果的に様々な感情を誘発することができると考えられるため、分類したいクラス数に応じて Math の難易度段階を変更し、また、仮説検証に基づいて妥当な実験参加者の人数を確保する必要がある。さらに、リラックス期間のサーモグラフィも撮影し、主観ストレス強度と同様にベースラインにより RVS を正規化することで精度が改善される可能性がある。また、心理実験によるデータ収集に加え、前処理とモデル性能についても改善が必要である。まず、前処理に関して、本研究では、先行研究 [2] と同様に、赤外線カメラによって撮影した鼻孔領域サーモグラフィから 1 次元の呼吸信号を復元し、短時間フーリエ変換を適用することで呼吸周波数を抽出して、深層学習に用いた。しかし、呼吸周波数のみでは、ストレスあるいは感情推定に必要となる特徴が少ない可能性があるため、呼吸位相等を前処理によって抽出することで精度向上が期待される。次に、モデル性能に関して、RVS は時間軸方向で呼吸周波数が変動する特徴であるため、時間軸方向の畳み込み処理可能な 2D-CNN 構成とすることで最適な学習ができる可能性がある。あるいは、

4.5 結言

RVS のみでは、分類精度の向上が困難であると考えられるため、RVS に加え、文献 [18] ように鼻先の温度変化や脈派等、複数の生体信号を用いる深層学習法について検討する必要がある。

4.5 結言

本章では、3 クラス分類時における “Low-level-stress” の F 値向上に向け、RVS に基づくストレス推定までの流れそれぞれに対する提案手法の評価を行った。まず、条件を統制した実験にて収集した参加者データの解析として、主観ストレスと各課題 (Math と Stroop) における 2 群間 (easy と difficult) の平均値の差について検定した結果を述べた。加えて、課題正答率および反応速度についても検定を行い、その結果に触れた。次に、提案手法を適用したストレスの推定精度についてそれぞれ単独の有効性を評価し、その後、有効性の高い組み合わせにて分類精度および F 値の評価を行った。結果として、本実験により収集したデータに対して、Baseline と Balanced(Oversampling) を適用すると先行研究データ [2] に Balanced(Oversampling) を適用した場合と比較して “Low-level-stress” の F 値が大幅に向上した。しかし、依然として、分類精度が改善されなかったため、今回、本研究にて提案した手法それぞれに対して更なる改善が必要であるため、本章の最後にて、分類精度向上へ向けた課題考察について言及した。

第 5 章

結論

現代社会は、経済的豊かさや科学技術の発展によって、より快適で利便性に満ち溢れている反面、ストレス社会と言われている [3]。特に、仕事上でストレスを抱える労働者は約 6 割を超え、日々のストレスの蓄積によって精神面のみならず身体面の病気が深刻化している [6]。そのため、今日のストレス社会において、自身のストレスの状態を正確に理解することや個人を取り巻く周囲の支援を充実させることは健康維持や病気の予防に重要な意味を持つ。それに伴い、近年では、心拍数や心拍変動、脳波といった生体信号を用いたストレス推定法について研究されている [1][11]。しかし、ストレスを推定する多くの手法では、生体信号の計測に接触型のセンサを用い、十分に統制された実験環境を必要とする等、計測に関わる制約が多い。一方、非接触型のセンサによるストレス推定法についても研究されており [12][13]、Cho らは鼻孔領域サーモグラフィから観測した呼吸パターン (RVS) に基づくストレス推定法を提案した [2]。この手法では、2 層 2D-CNN による 2 クラス分類時の精度が 84.6% で、3 クラス分類時の精度が 56.5% と報告され、日常生活の中で非侵襲的かつ簡易にストレス推定する実現可能性が示唆された。ただ、2 クラス分類時における “None-stress” および 3 クラス分類時における “Low-level-stress” の F 値がチャンスレベル付近であり、十分に高い精度ではない。

そこで、本稿の第 2 章では、まず初めに RVS に基づくストレス推定の流れについて述べ、その後、先行研究 [2] で使用されていた同一データセットを用いて、2 層 2D-CNN [2] の他、多層 2D-CNN と ImageNet に基づく転移学習によって 2 クラス分類と 3 クラス分類の追検証を実施した。その結果、2 クラス分類時は、転移学習によって精度が 10%

程度向上したが、3 クラス分類時は、いずれも向上せず、また、転移学習を適用した場合、3 クラス分類時における“Low-level-stress”の F 値が数%と極端に低い値であることを確認した。3 クラス分類時における精度が改善されなかった原因として、主にデータセットの質とモデル性能が考えられ、前者に関しては、RVS と主観ストレス強度の個人差が大きく、また、実験参加者毎に各ストレスクラスのラベル数に偏りがあるため全てのストレス強度を均等に学習できていないと考えられる。また、後者に関して、RVS は“None-stress”や“High-level-stress”に比べて“Low-level-stress”固有の特徴を含んでいないため、多層の畳み込みでは十分に学習できていないと考えられる。これらの課題を踏まえ、本研究では、RVS に基づくストレス推定における 3 クラス分類時の精度向上を目的とした。

第 3 章では、データセットの質とモデル性能の改善点に対して RVS に基づくストレス推定までの 3 つの流れに基づいて改善策について述べた。まず、心理実験によるデータ収集では、個人差の抑制およびラベルの均等化に向けて先行研究 [2] の実験手続きを基に条件を統制した。次に、前処理およびデータセット構築では、RVS と主観ストレス強度の個人差を緩和させる手法とラベルの均等化、さらに、課題別データセットの分割手法を検討した。そして、深層学習によるストレス推定の改善手法として、データ拡張および先行研究データ [2] を基に既存モデルをファインチューニングする事前学習タスクの改善によって更なる特徴抽出能力の向上を試みた。

第 4 章では、第 3 章にて提案した改善手法について先行研究データ [2] と本実験により収集したデータを対象に評価を行った。結果として、本実験データを対象に、ベースラインによる主観ストレス強度の正規化 (Baseline) および不均等なデータセットの均等化 (Balanced) をすることで先行研究データ [2] に Balanced のみ適用した場合と比較して、“Low-level-stress”の F 値が最大で 45%程度向上することを確認した。この結果より、提案手法では、主観ストレス強度の個人差が緩和されたことで“Low-level-stress”固有の特徴を学習できていたと考えられる。その一方、精度に関しては、期待する結果が得られなかった。そのため、今後は、本実験データに対して Baseline と Balanced を適用

した“Low-level-stress”のF値を維持したまま精度の向上が必要である。ストレスの他、疲労度、快不快、覚醒眠気の推定結果より、RVSに基づく感情推定の精度は、心理実験にて収集するデータの質に大きく依存すると考えられるため、今後の課題として、更なる条件を統制した心理実験にてデータを収集し、それに加え、呼吸周波数の他、呼吸位相等の特徴を前処理によって抽出することで精度の改善が期待される。あるいは、RVSのみでは、分類精度の向上が不十分と考えられるため、複数の生体信号を用いる深層学習法について検討する必要がある。

今後、現実的に簡易なストレス推定法が確立されることで、日々のストレス状態を個人が正確に理解し、健康維持や病気の予防に役立つのではないだろうか？ 実際に、スマートフォン等で簡便にストレス強度を計測し、自分自身で確認可能になることで、ストレス緩和のための呼吸法や音楽、匂い等のリアルタイムフィードバックによるリラックス効果に貢献できるのではないかと考えられる。また、日々のストレス強度を収集し、その履歴を心理カウンセラーや心療内科医師に提示することで、個人に合ったストレスに対する自己訓練法や生活習慣や環境の見直し等、より効果的な治療法の検討に有用なデータになると考えられる。

謝辞

本研究を遂行するにあたり、日頃より非常に多くのご指導とご助言を賜りました指導教員の岩田誠教授には心より深く感謝申し上げます。懇切丁寧なご指導頂きましたことで本論文を完成させることができました。

本研究の副査をお引き受けて下さり、貴重なご質問とご助言を賜りました繁桝博昭教授、並びに中原教授に感謝申し上げます。

実験の遂行にあたり、ご助言を賜りました渡邊言也助教に感謝申し上げます。また、データを提供して頂きましたロンドン大学の Youngjun Cho 先生に御礼申し上げます。

快く、実験に参加頂きました皆様、並びにご協力頂きました皆様には御礼申し上げます。

研究室の先輩として日頃よりご指導・ご協力頂きました、博士課程 2 年の Zhen Zhang 氏、博士課程 1 年の Valeeprakhon Tamnuwat 氏には心より感謝申し上げます。特に、Zhang 氏には、研究方針について時間を惜しむことなく質問や相談に乗って頂き、全面的に協力頂きましたこと心より感謝いたします。

研究室の同輩である岡野秀平氏、尾ノ井嶺卓氏、古田雄大氏には、日頃より励み合うことで幾度となく支えられました。ありがとうございました。

また、研究室の後輩として幅広くご支援して頂きました、高橋龍一氏、植本陸氏、坂口白磨氏、松坂拓海氏、計屋和希氏、石橋璃貴氏、伊藤雅俊氏、市ノ木一希氏、岡村健勝氏、椎葉啓介氏、山下拓巳氏には深く感謝いたします。

最後になりましたが、ここまで私に学ぶ機会を与えて下さり、また、日頃より多方面で支えて頂きました家族には、心より感謝申し上げます。これまでの貴重な経験を糧に日々精進してまいります。本当にありがとうございました。

参考文献

- [1] J. Healey and R. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” IEEE Trans. Intelligent Transportation Systems, Vol. 6, No. 2, pp. 156–166, June, 2005.
- [2] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, “DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings,” In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction, ACII, pp. 456–463, 2017.
- [3] 杉浦こころのクリック., “ストレス社会がつくるこころの病”. <https://sugiura-kokoro.com/stress-2.html>. 参照 Jan.31, 2023.
- [4] 厚生労働省., “ストレスとは”. <https://kokoro.mhlw.go.jp/attentive/atv001/>. 参照 Jan.31, 2023.
- [5] 厚生労働省., “令和 3 年「労働安全衛生調査（実態調査）」の概況”. <https://www.mhlw.go.jp/toukei/list/r03-46-50b.html.pdf>. 参照 Jan.31, 2023.
- [6] 品川メンタルクリニック., “ストレスと関連する主な障害について”. <https://www.shinagawa-mental.com/stresscare/>. 参照 Jan.31, 2023.
- [7] 厚生労働省., “休養・こころの健康”. https://www.mhlw.go.jp/www1/topics/kenko21_11/b3.html. 参照 Feb.1, 2023.
- [8] 石原俊一, “ストレスとその影響”, 心理学概論, 山内弘継・橋本宰監修, 岡市廣成・鈴木直人編集, 2009, ナカニシヤ出版, pp. 312–320.
- [9] H. Selye, “現代社会とストレス”, 杉靖三郎・田多井吉之助訳, 1988, 法政大学出版局.
- [10] R. S. Lazarus, S. Folkman, “ストレスの心理学—認知的評価と対処の研究”, 本明寛・春木豊・織田正美監訳, 1991, 実務教育出版.
- [11] M. Rimini-Doering, D. Manstetten, T. Altmueller, U. Ladstaetter, and M. Mahler,

- “Monitoring driver drowsiness and stress in a driving simulator,” First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 58–63, 2001.
- [12] H. Kataoka, H. Yoshida, A. Saijo, M. Yasuda, and M. Osumi, “Development of a skin temperature measuring system for Non-contact stress evaluation,” In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, pp. 940–943, Aug., 1998.
- [13] V. Engert, A. Merla, J. A. Grant, D. Cardone, A. Tusche, and T. Singer, “Exploring the Use of Thermal Infrared Imaging in Human Stress Research,” PLOS ONE, Vol. 9, No. 3, pp. e90782, Mar. 2014.
- [14] Y. Cho, S. J. Julier, N. Marquardt, and N. Bianchi-Berthouze, “Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging,” Biomedical optics express, 8(10), pp. 4480–4503, 2017.
- [15] H. Kaiming, Z. Xiangyu, R. Shaoqing and S. Jian, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385, Dec., 2015.
- [16] H. Andrew, S. Mark, C. Grace, C. Liang-Chieh, C. Bo, T. Mingxing, W. Weijun, Z. Yukun, P. Ruoming, V. Vijay, V.L. Quoc and A. Hartwig, “Searching for MobileNetV3,” arXiv:1905.02244, Nov. 2019.
- [17] T. Mingxing and V.L. Quoc, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” arXiv:1905.11946, Sep. 2020.
- [18] Y.Cho, S. J. Julier and N. Bianchi-Berthouze, “Instant Stress: Detection of Perceived Mental Stress Through Smartphone Photoplethysmography and Thermal Imaging,” JMIR Ment Health, Vol. 6, No. 4, pp. e10140, Apr., 2019.