

# 年齢認識および性別認識における 手法・データセット・クラス不均衡に起因する 精度変化の要因調査

---

尾崎 匠, 西山 正志 (鳥取大)

# 研究背景

- 年齢や性別の属性認識はマーケティング・防犯システムなどで活用される重要なタスク
- 属性認識は視覚言語モデル (VLM) を導入した手法※<sub>1,2,3</sub> により精度が向上
- 属性認識では様々な要因が精度に影響を与える

## 精度変化の主な要因

### □ 手法



- VLM
- ViT
- CNN

### □ データセット



- 撮影ロケーション
- 被写体の多様性
- 画像枚数

### □ クラス不均衡



- 対策なし
- ランダムサンプリング
- 損失関数

対策

※データセットにおけるクラス不均衡

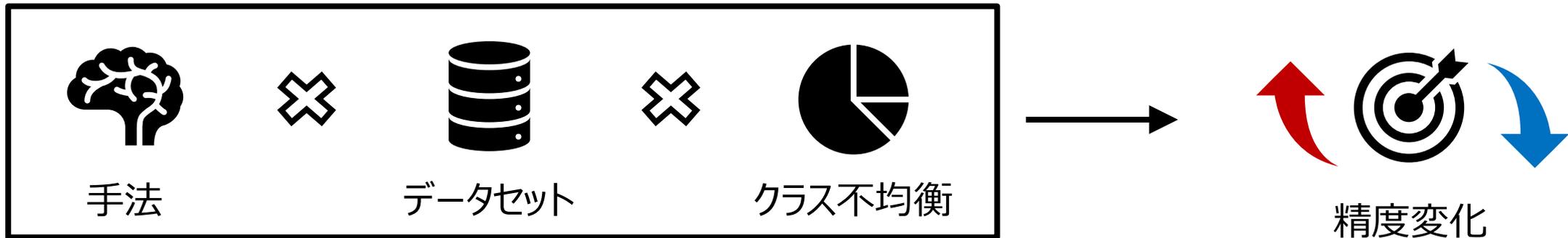
- 属性クラス間のサンプル数の偏り
- 属性クラスの付け方

これらの要因が互いに絡み合い精度変化にどのように影響を及ぼすか明らかでない

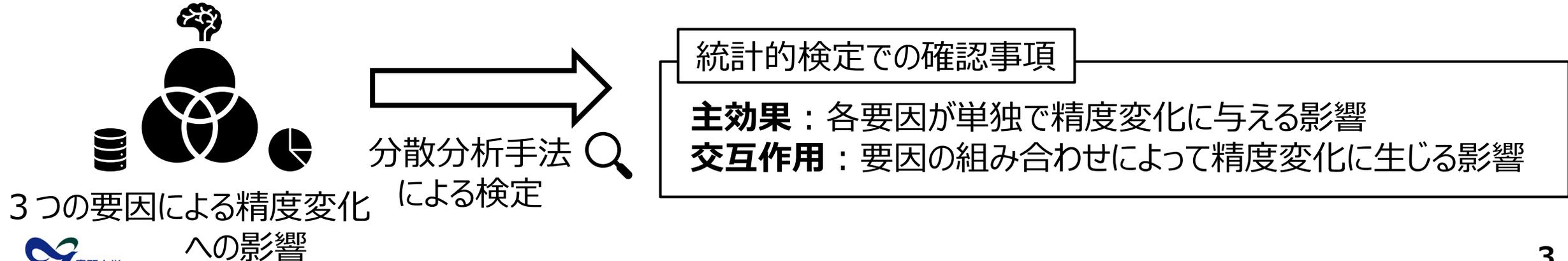
# 研究目的

## 年齢・性別の属性認識において手法・データセット・クラス不均が精度変化に及ぼす影響を統計的に検証

- 手法・データセット・クラス不均衡は独立ではなく相互に影響し合う



- 多要因分散分析で3つの要因による精度変化の**主効果**と**交互作用**を検定する



# 仮説と各要因の傾向

仮説：年齢・性別の属性認識において手法・データセット・クラス不均衡のいずれの要因も精度変化に強く影響する

## 年齢・性別を認識する際の各要因の傾向



手法

VLM, ViT, CNNなどモデルアーキテクチャの異なる手法が存在



データセット

データセットによって撮影環境・被写体の多様性・画像枚数が異なる

※データセットにおける  
クラス不均衡の傾向

年齢：若年層・高齢層の画像枚数が少ない傾向  
性別：女性の画像枚数が少ない傾向



クラス不均衡

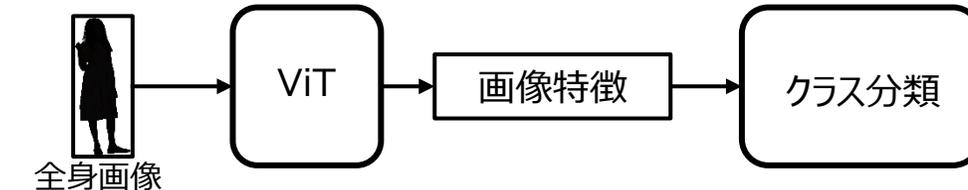
クラス不均衡の対策にいくつか手法が存在

# 要因：手法 (8水準：M1-M8)

ViT・CNN・VLM等のモデルアーキテクチャの異なる複数の属性認識手法を用いる

M1: PARFormer [F. Xinwen et al., TCSVT (2023)]

- ViTを使用
- 属性認識に有効な特徴を抽出

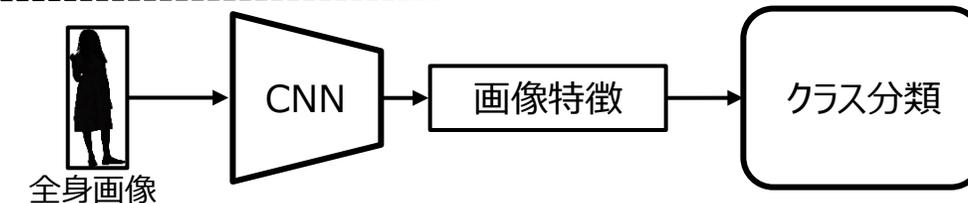


M3: CLIP-Adapter [G.Peng et al., IJCV (2024)]

[G.Peng et al., IJCV (2024)]

M2: SSC [J.Jia et al., ICCV (2021)]

- CNNを使用
- 属性認識に有効な領域・特徴を抽出

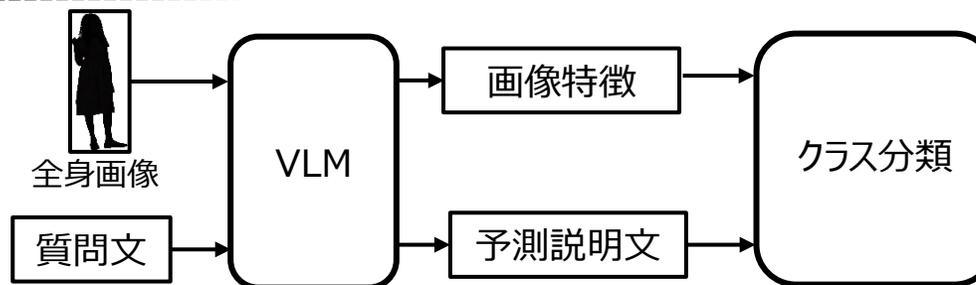


M4: CoOp [Z.Kaiyang et al., IJCV (2022)]

[Z.Kaiyang et al., IJCV (2022)]

M7: LLMPAR [J.Jin et al., arXiv:2408.09720 (2024)]

- 画像特徴, テキスト特徴を使用
- VQAによって生成された文章を認識に使用



M5: Visual-Prompt [H.Bahng et al., arXiv:2203.17274 (2022)]

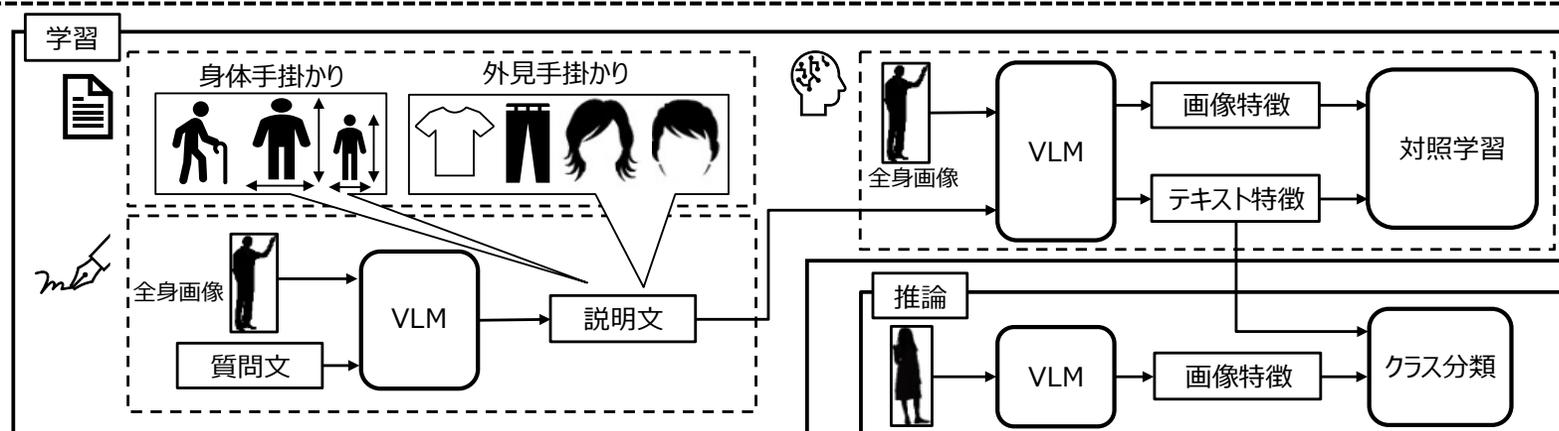
[H.Bahng et al., arXiv:2203.17274 (2022)]

M6: PromptPAR [X.Wang et al., TCSVT(2024)]

[X.Wang et al., TCSVT(2024)]

M8: Ours [T.Ozaki et al., IEEE Access (2025)]

- 画像特徴, テキスト特徴を使用
- VQAで生成した詳細な説明文を用いて抽出したテキスト特徴を対照学習に用いる



# 要因：データセット (4水準：D1-D4)

撮影環境・被写体の多様性・画像枚数の異なる複数のデータセットを用いる

## D1: RAPv2.0

[L.Dangwei et al., IEEE TIP (2018)]

- 屋内ショッピングモールで撮影
- 買い物客特有の特徴（荷物等）
- 画像枚数：84,420枚



## D2: PETA

[D.Yubin et al., ACM (2014)]

- 複数データセットを統合しており、様々な撮影環境を含む
- カジュアルな服装の一般的な歩行者
- 画像枚数：18,972枚



## D3: MSP60k

[J.Jin et al., arXiv:2408.09720 (2024)]

- 学校、建設現場等多様なシーン
- 学校の生徒・作業環境の服装などを含む
- 画像枚数：60,122枚



## D4: PA100k

[X.Liu et al., ICCV(2017)]

- 屋外環境が中心
- カジュアルな服装の一般的な歩行者
- 画像枚数：100,000枚



# データセットにおけるクラス不均衡の具体例

- 若年層・高齢層の画像枚数が少ない傾向
- データセットによって年齢クラスの区切りが異なる

D1: RAPv2.0		D2: PETA		D3: MSP60k		D4: PA100k	
年齢クラス	画像枚数	年齢クラス	画像枚数	年齢クラス	画像枚数	年齢クラス	画像枚数
≤16	750	≤15	170	child	6767	<18	1469
17-30	34080	16-30	9443	adult	51511	18-60	92844
31-45	46454	31-45	6246	elderly	1844	>60	5687
46-60	2946	46-60	1943	-	-	-	-
>60	190	>60	1170	-	-	-	-

- 女性の画像枚数が少ない傾向

性別クラス	D1: RAPv2.0	D2: PETA	D3: MSP60k	D4: PA100k
男性	58015	10415	36045	54664
女性	26405	8557	24077	45336

**年齢・性別ともにクラス不均衡が存在**

# 要因：クラス不均衡 (3水準：I1-I3)

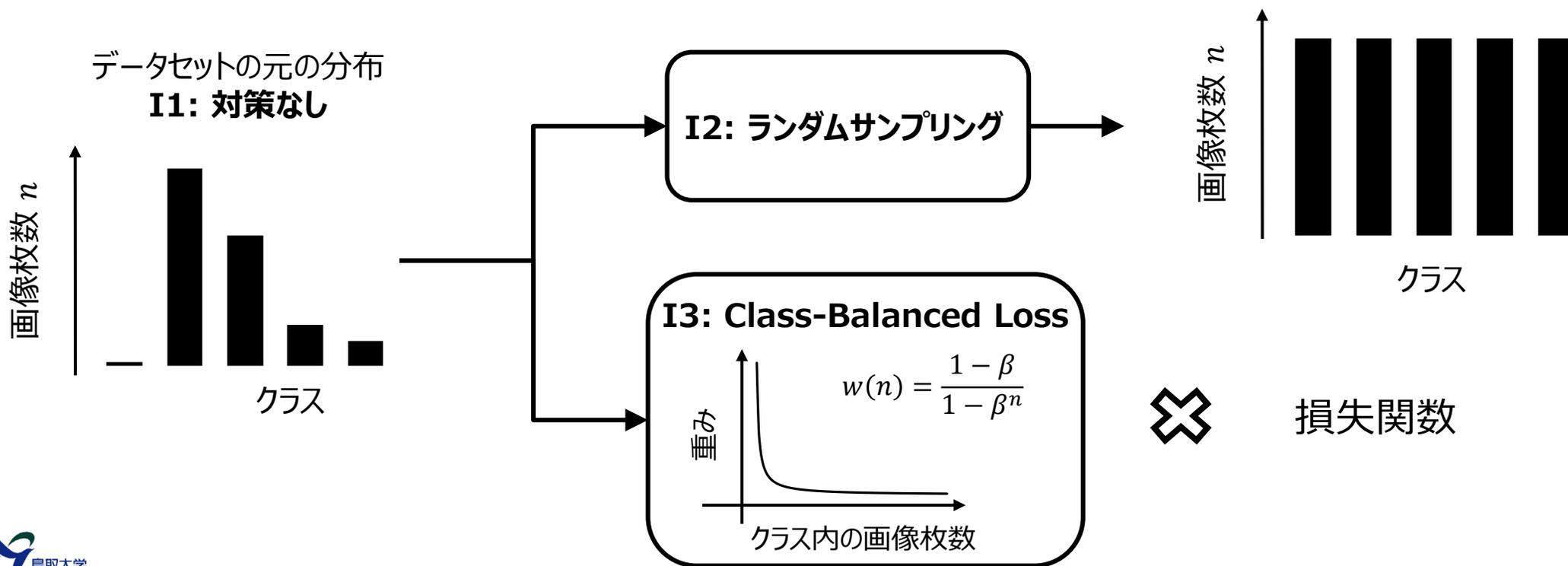
クラス不均衡の対策手法を複数用いる

I1: 対策なし・・・データセットの元の分布を維持

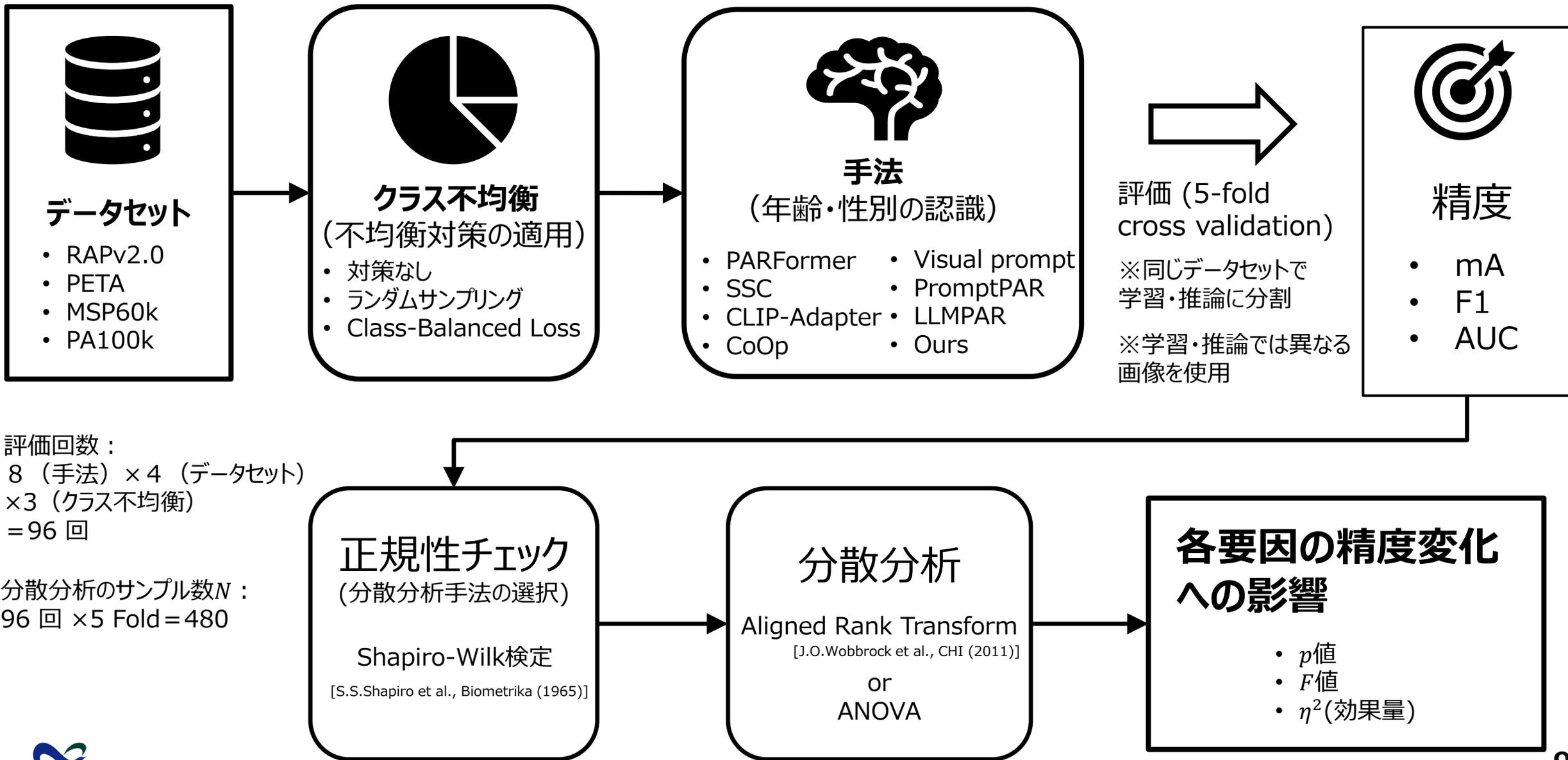
I2: ランダムサンプリング・・・少数クラスを多数クラスに枚数を合わせるようにアップサンプル  
データセット間でも枚数を揃えるためランダムサンプル

I3: Class-Balanced Loss・・・画像枚数に応じて損失関数に重みを掛ける手法

[Y.Cui et al., CVPR (2019)]



# 実験デザイン



# 年齢を認識する際の各要因の影響

手法・データセット・クラス不均衡の主効果・交互作用をAligned Rank Transformで分析

\*  $p < 0.001$

$\eta^2$ は効果量

要因	自由度	F1		mA		AUC	
		$\eta^2$	F値	$\eta^2$	F値	$\eta^2$	F値
<b>主効果</b>							
手法 * 	7	<b>0.95</b>	996	<b>0.92</b>	601	0.82	248
データセット * 	3	0.88	966	0.89	1016	<b>0.85</b>	746
クラス不均衡 * 	2	0.68	406	0.77	626	0.09	19
<b>交互作用</b>							
手法 × データセット *	21	<b>0.85</b>	99	<b>0.83</b>	88	<b>0.82</b>	83
手法 × クラス不均衡 *	14	0.62	45	0.80	108	0.71	68
データセット × クラス不均衡 *	6	0.50	62	0.57	84	0.15	11
手法 × データセット × クラス不均衡 *	42	0.66	18	0.60	14	0.59	13

- 手法・データセット・クラス不均衡の**主効果および交互作用が有意**に確認され全要因が精度変化に強く影響するという仮説が支持された
- 特に**手法とデータセット**の主効果・交互作用が大きい

# 性別を認識する際の各要因の影響

手法・データセット・クラス不均衡の主効果・交互作用をAligned Rank Transformで分析

\*  $p < 0.001$

$\eta^2$ は効果量

要因	自由度	F1		mA		AUC	
		$\eta^2$	F値	$\eta^2$	F値	$\eta^2$	F値
<b>主効果</b>							
手法 * 	7	<b>0.97</b>	1174	<b>0.97</b>	1299	<b>0.96</b>	774
データセット * 	3	0.90	799	0.91	813	0.89	655
クラス不均衡 *  ※Class-Balanced Loss を除いた2水準	1	0.67	504	0.59	368	0.53	282
<b>交互作用</b>							
手法 × データセット *	21	<b>0.89</b>	97	<b>0.89</b>	101	<b>0.90</b>	113
手法 × クラス不均衡 *	7	0.61	55	0.54	42	0.19	8
データセット × クラス不均衡 *	3	0.43	64	0.44	65	0.46	70
手法 × データセット × クラス不均衡 *	21	0.44	9	0.40	8	0.29	5

- 手法・データセット・クラス不均衡の**主効果および交互作用が有意**に確認され全要因が精度変化に強く影響するという仮説が支持された
- **手法**の主効果が最も大きく、**手法とデータセット**の交互作用が大きい

# 【ご参考】手法に注目した時の属性認識の平均精度 [%]

手法	年齢			性別		
	F1	mA	AUC	F1	mA	AUC
PARFormer [F. Xinwen et al., TCSVT (2023)]	72.7	70.0	87.7	91.6	91.5	96.4
SSC [J.Jia et al., ICCV (2021)]	72.8	70.0	90.0	92.2	92.1	97.2
CLIP-Adapter [G.Peng et al., IJCV (2024)]	69.5	67.5	89.1	86.7	86.6	92.9
CoOp [Z.Kaiyang et al., IJCV (2022)]	65.3	65.8	89.8	84.5	84.3	92.0
Visual-Prompt [H.Bahng et al., arXiv:2203.17274 (2022)]	50.7	50.4	78.8	85.8	85.7	90.4
PromptPAR [X.Wang et al., TCSVT(2024)]	75.1	73.1	<b>92.0</b>	94.2	94.2	96.8
LLMPAR [J.Jin et al., arXiv:2408.09720 (2024)]	52.1	65.3	87.6	94.6	94.8	<b>98.4</b>
Ours [T.Ozaki et al., IEEE Access (2025)]	<b>78.8</b>	<b>77.8</b>	91.9	<b>95.3</b>	<b>95.3</b>	98.3

提案手法を含むVLMを使用した手法が全体として高い精度を示す傾向が見られた

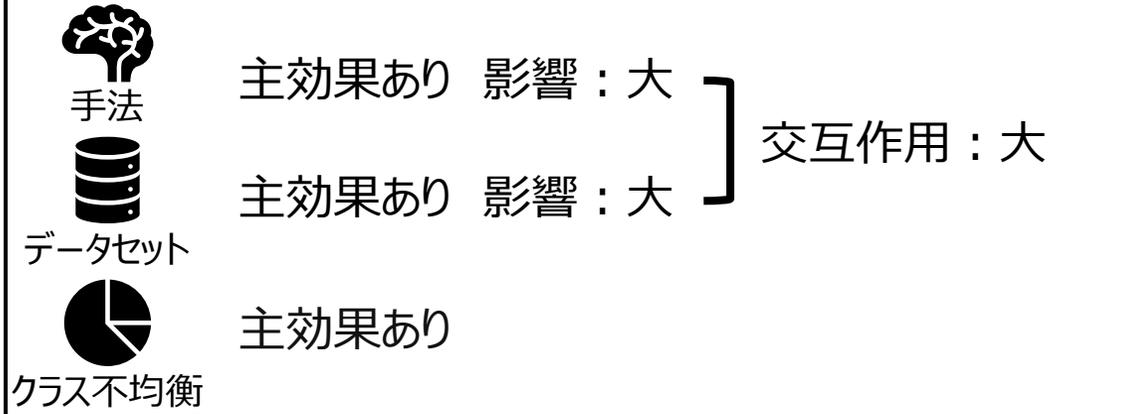
# まとめ

## 年齢認識および性別認識において手法・データセット・クラス不均衡が 精度変化に及ぼす影響を統計的に検証

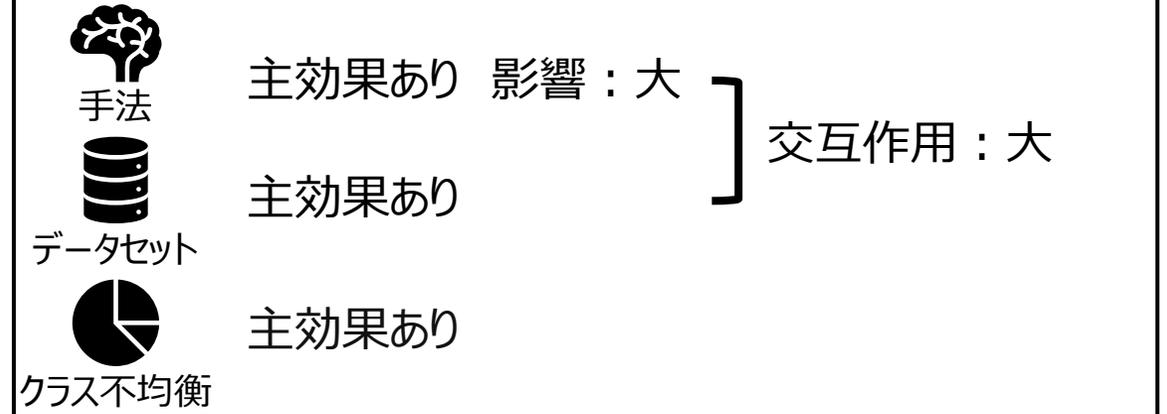
**仮説**：年齢・性別の属性認識において手法・データセット・クラス不均衡のいずれの要因も精度変化に強く影響する

**結果**：いずれの要因も精度変化に強く影響した。特に手法・データセットが強く影響した

### 年齢を認識する際の各要因の精度変化への影響



### 性別を認識する際の各要因の精度変化への影響



### 今後の課題

- クラス不均衡対策の効果量が手法・データセットに比べて低かった原因の追求
- 年齢認識・性別認識以外の属性への展開