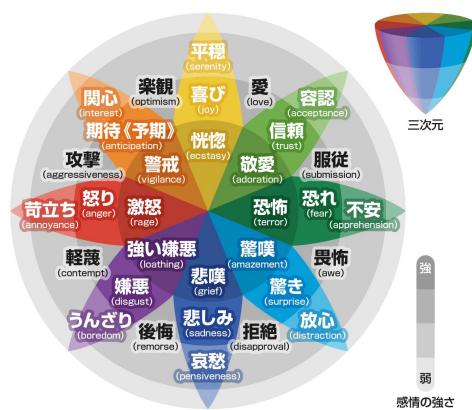


# 乳児の表情や仕草からその感情はわかるのか？

## 分析の背景と目的

- ころころと変わる乳児の感情を理解するのは困難(パパ・ママは大変)。
- アプリケーションによって、乳児の感情が理解できるようになれば、育児がもっと楽になるのではないか？
- プルチック理論\*の8つの基本感情を基に、乳児の表情や仕草を10分類し、判定モデルを作成し、アプリ開発のヒントとする。



\*「プルチックの感情の輪 | 人間の感情は色で分類すると関連性がわかる？」  
<https://swingroot.com/plutchik-emotion/>

画像分類	定義	感情分類
あはは (ahaha)	満たされて穏やかに笑っている様子。口角を上げて手足をバタバタさせたりしている。	喜び
びくびく (bikubiku)	突然の物音や見慣れない人に対して、不安げな目をして、じっと身構えている様子。	不安
ぼー (boo)	口を半開きにしてぽかんとあらぬ方を見つめる様子。特に、左上を見つめる時に出る表情。	放心
ぐずぐず (guzuguzu)	思い通りならず、くしゃくしゃの表情でいまにも泣きそうな状態。ミルクを待つ時に多い。	苛立ち
はむはむ (hamham)	ミルクやおしゃぶり、自分の手を咥えている様子。口元に物体があることが多い。	恍惚
いやいや (iyaiya)	眉間にしわを寄せて、顔をそむけて拗ねたような様子。ただし泣いてはいない。	嫌悪
じー (jii)	目を大きく開いて、あちこちを見回しながら新しいものに興味津々の様子。	関心
にこにこ (nikoniko)	警戒心が解け、リラックスしている様子。言葉通り、にこにこしている状態。	容認
すやすや (suyasuya)	安心してぐっすりと寝息を立てている様子。	平穡
うわーん (uwan)	物事がうまくいかずに、口を開けて泣いている様子。いやいやから発展することが多い。	悲しみ

## 混同行列の3モデル比較

\*分析対象は小島次女(生後約2ヶ月)で、各表情を50枚ずつ撮影して分類。テストデータは各10枚、合計100枚。

ahaha



bikubiku



boo



guzu



hamham



iyaiya



iii



nikoniko



suyasuya

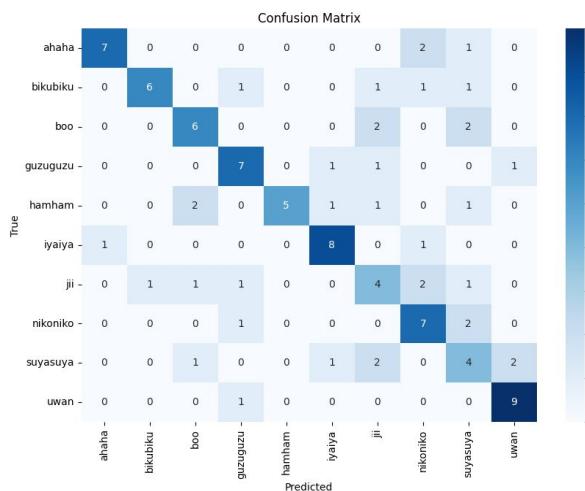


uwan

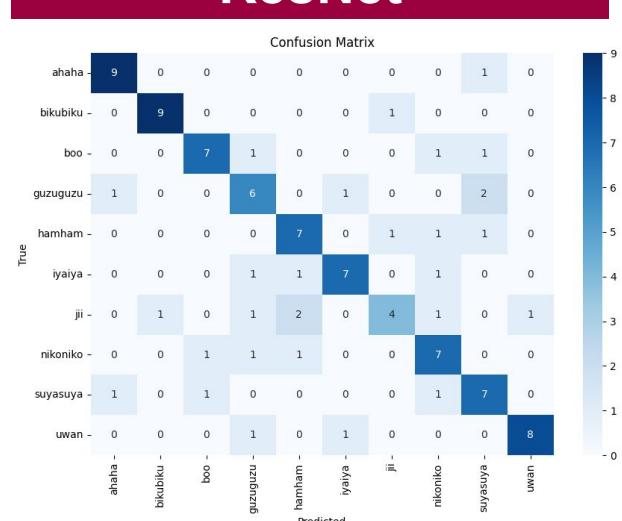


#### ・モデル別の混同行列

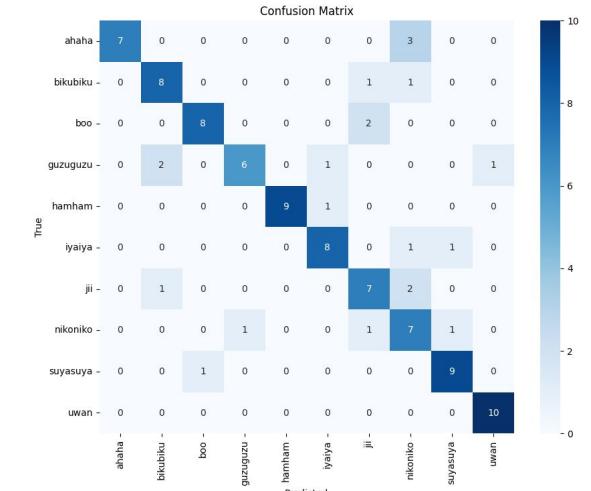
CNN



# ResNet

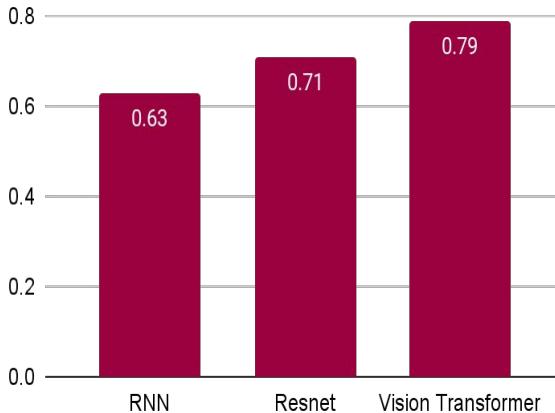


# Vision Transformer



# 予測結果値の比較

## ・モデル別の accuracy



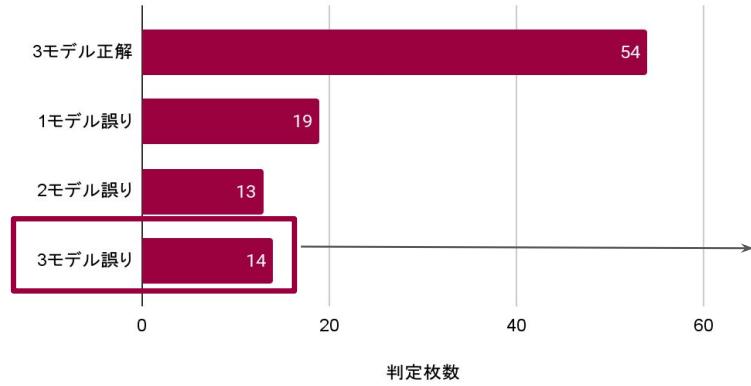
## ・モデル別の精度比較

分類／スコア	CNN			ResNet			Vision Transformer		
	precision	recall	fi-score	precision	recall	fi-score	precision	recall	fi-score
あはは	0.88	0.7	0.78	0.82	0.9	0.86	1	0.7	0.82
ぴくぴく	0.86	0.6	0.71	0.9	0.9	0.9	0.73	0.8	0.76
ぼー	0.6	0.6	0.6	0.78	0.7	0.74	0.89	0.6	0.71
ぐずぐず	0.64	0.7	0.67	0.55	0.6	0.57	0.86	0.6	0.71
はむはむ	1	0.5	0.67	0.64	0.7	0.67	1	0.9	0.95
いやいや	0.73	0.8	0.76	0.78	0.7	0.74	0.8	0.8	0.8
じー	0.36	0.4	0.38	0.67	0.4	0.5	0.64	0.7	0.67
にこにこ	0.54	0.7	0.61	0.58	0.7	0.64	0.5	0.7	0.58
すやすや	0.33	0.4	0.36	0.58	0.7	0.64	0.82	0.9	0.86
うわーん	0.75	0.9	0.82	0.89	0.8	0.84	0.91	1	0.95

- 3モデルの正解率は0.63-0.79
- CNNとResNetはぐずぐずとじーの混同が多い一方、ViTはにこにことじーの混同が目立つ。
- 全モデルでぐずぐずといやいやの誤分類が共通していた。
- すやすやはCNNからViTで大幅改善

# 考察

## ・画像別に見たモデル判定状況



- 3モデル全てで判別を誤った画像が 14枚あったが、自分自身のラベリングが正しいか疑問に残る結果となった (No5、6、8)。人の判断が全て正しいわけではなく、AIの判定から再びラベリングすることも重要ではないか。
- 全画像の確率差の平均は約 0.46であったのに対して、13件のケースでは確率差が 0.2未満と小さく判別の難しさが伺えた。上位ラベルの確率が複数、同時に出来ることで解釈に幅が出て有効だろう。
- 乳児の表情は刻一刻と変わるために、時系列モデルを活用することで表情の変化のパターンから、より正確にその感情を捉えることができるかもしれない。

## ・3モデル全てで判別を誤った主な画像



1. True: じー

CNN: ぐずぐず  
ResNet: ぐずぐず  
ViT: にこにこ



2. True: ぐずぐず

CNN: じー  
ResNet: すやすや  
ViT: びくびく



3. True: じー

CNN: びくびく  
ResNet: うわーん  
ViT: びくびく



4. True: いやいや

CNN: にこにこ  
ResNet: はむはむ  
ViT: すやすや



5. True: にこにこ

CNN: ぐずぐず  
ResNet: ぐずぐず  
ViT: ぐずぐず



6. True: じー

CNN: にこにこ  
ResNet: にこにこ  
ViT: にこにこ



7. True: にこにこ

CNN: すやすや  
ResNet: ぼー  
ViT: すやすや



8. True: ぐずぐず

CNN: いやいや  
ResNet: いやいや  
ViT: いやいや