Transformerモデルを用いた胎児心拍数陣痛図によるハイリスク出産の予測

辺見, 一成 / HEMMI, Kazunari

(出版者 / Publisher)
法政大学大学院理工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学研究科編
(巻 / Volume)
65
(開始ページ / Start Page)
1
(終了ページ / End Page)
7
(発行年 / Year)
2024-03-24
(URL)
https://doi.org/10.15002/00030779

Transformer モデルを用いた 胎児心拍数陣痛図によるハイリスク出産の予測

PREDICTION OF HIGH-RISK BIRTHS FROM CARDIOTOCOGRAM DATA USING TRANSFORMER-BASED MODELS

辺見一成
Kazunari HEMMI
指導教員　柴田千尋

法政大学大学院理工学研究科システム理工学専攻修士課程

The purpose of the study is to explore effective methods on Transformer-based models for Cardiotocogram (CTG). In this study, we use clinical data which is collected from multiple medical institutions. We conduct a certain number of experiments to verify the effectiveness of various recent model architectures. In addition, the data augmentation methods and the specific types of signal processing are examined. Both CutMix and Slit-type CutMix (the proposed method), which are data augmentation methods, are shown to be effective through experiments. We especially demonstrate that the latter is remarkably helpful in order to improve prediction accuracy for newly introduced test datasets.

Key Words : *Deep Learning, Signal Processing, Medical Application of AI*

1. はじめに

一般的に，臨床において，産婦人科医は子宮内の胎児が直面するリスクを胎児心拍数陣痛図 (Cardiotocogram : CTG) から目視で判断している．CTG は周産期医療において胎児の状態を与えてくれる唯一のツールであるが，検査者間の診断の再現性が低いことが問題とされている．また，CTG は多くのノイズが含まれており，診断の妨げとなっている [1].

本研究では，Transformer [2] をベースとした深層学習モデルである VisionTransformer (ViT) [3], PoolFormer [4][5], Swin Transformer [6] を用いて CTG から重要な特徴を取り出し，ハイリスクな出産を予測することを目的とする．また，様々な実験を通してモデル構造や波形の前処理，データ拡張を検証することで時系列データである CTG に対して有効な手法を探索する．

2. 胎児心拍数陣痛図

胎児心拍数陣痛図 (CTG) は分娩監視装置による胎児心拍数 (Fetal Heart Rate : FHR) と子宮収縮 (Uterine Contraction : UC) の連続記録である．FHR と UC は別のグラフとして上段と下段に記録されており，単位は bpm と mmHg である (図 1). 産婦人科医は UC に対して FHR の変化を見ることで胎児の健常性を推測する [1].

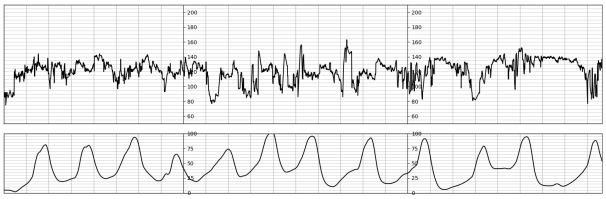

図 1 胎児心拍数陣痛図

3. 先行研究

Zhao [7] らは 8 個のレイヤーで構成される Convolutional Neural Network (CNN) モデルを提案し，FHR を CWT 変換した画像を用いて胎児酸欠症を予測している．彼らは FHR から時間領域と周波数領域の両方に隠れた特徴量が重要だとしている．また，Anwar [8] らは深層学習モデルを改良する手法を提案し，低酸素症を予測している．彼らはデータの前処理やモデルの層を増加させることで予測精度を向上させている．さらに，Alkanan [9] らは EfficientNet [10] をベースにしたマルチインプットモデルを提案し，ハイリスクな出産を予測している．彼らのモデルは CTG に加えて妊娠期間数を入力として扱うことが重要だと考え，２つの要素を同時にマルチインプットモデルに入力する手法を提案している．

4. データセット

本研究では複数の医療機関から得た 37,649 件のデータセットのうち，ノイズ処理を行なった結果得られた 34,266 件を使用する．使用するデータセットの詳細を表 1 に示す．

表 1 CTG データセット

医療機関名	タイプ	データセット	サンプル数
九州大学病院	大学病院	学習用	2,486
熊本大学病院	大学病院	学習用	1,044
福岡大学 (-2018)	大学病院	学習用	1,828
福田病院	個人病院	学習用	19,360
井槌病院	個人病院	学習用	4,457
東野産婦人科	個人病院	学習用	3,909
福岡大学 (2018-)	大学病院	テスト用	896
ながかわ産婦人科	個人病院	テスト用	286
合計			34,266

(1) メタデータ

CTG には，母体年齢や妊娠期間数などのメタデータがセットで提供されている．メタデータの例を表 2 に示す．本研究では，メタデータのうち，それぞれ分娩後 1 分後と 5 分後の新生児の健常性を評価した値である，Apgar スコア 1 と Apgar スコア 5 から，ターゲットラベルを作成し，ハイリスクな出産を予測する．

表 2 メタデータ

データ項目	値
施設 ID	FU212335
ID	7712234
分娩年月日	2018/5/15
分娩時刻 [時]	8
分娩時刻 [分]	25
分娩様式	自然分娩
母体年齢	27
妊娠期間数	40
出生体重 [g]	3,363
性別	男
pH	7.318
Apgar スコア 1	7
Apgar スコア 5	8
退院時の生死	生

(2) 学習時に使用する画像の前処理

ノイズ処理を行った FHR と UC，10 分間の平均心拍数である胎児心拍数基線 (FHR baseline) をチャネル方向に連結させた画像を学習と検証に使用する．以下にノイズ処理と画像の作成について記述する．

a) ノイズ処理

CTG に含まれているノイズを除去するために FHR と UC に対して Alkanan ら [9] が用いたノイズ処理を行う．以下にノイズ処理の概要を示す．

- FHR に対するノイズ処理
  1. 0 が 15 秒以上続いた場合，欠損値 (NaN) とする
  2. 200bpm 以上，50bpm 以下の値は，欠損値とする
  3. 隣接する要素と比べて差が 25bpm 以上あった場合，欠損値とする
  4. 15 秒ごとの平均値と比べて差が 25bpm 以上あった場合，欠損値とする
  5. 全ての欠損値を補完する
- UC に対するノイズ処理
  1. 0 未満の場合，欠損値として補完する
  2. 60 秒の移動平均を取ったあと，指数平滑移動平均を取る

b) 学習に使用する画像の作成方法

30 分間の FHR と UC，FHR baseline をチャネル方向に連結させることで学習に使用する画像を作成する．CTG を構成する FHR と UC，FHR baseline のモノクロ画像を作成し，チャネル方向に連結させることで 3 チャネルの画像を作成する (図 2).

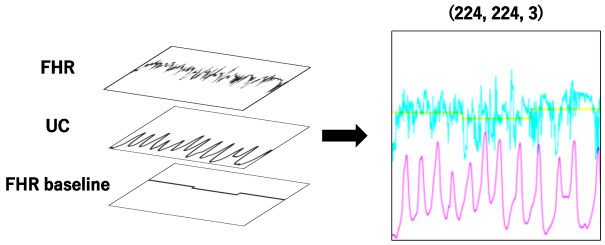

図 2 学習に使用する画像の作成過程

(3) Under-Sampling & Bagging

本研究で扱うデータセットはラベル比率が不均衡であるため，それを解消するための手法として，Under-Sampling と Bagging を用いる．データセットのターゲットラベルは 1 分後の Apgar スコアと 5 分後の Apgar スコアのどちらかが 6 以下であれば正例，それ以外は負例とする．作成したターゲットラベルの比率を表 3 に示す．

表 3 ラベル比率

データセット	正例	負例
学習用	739	32,187
テスト用	40	711

表 3 より，学習用データセットは正例に比べて負例は約 44 倍，テスト用データセットでは約 18 倍あり，ラベルが不均衡である．モデルを学習する際にデータセットのラベルが不均衡であるとモデルの予測結果が負例に偏ってしまう可能性があるため，学習用データセットの負例に対して Under-Sampling を行う．本稿では 10-fold 交差検証を行うため，各 fold ごとに Under-Sampling を適用する． 1 つ目の fold の学習用データセットに対して Under-Sampling を適用した場合のラベル比率を表 4 に示す．

表 4 Under-Sampling を行ったラベル比率

データセット	正例	負例
1 つ目の fold	648	28,968
1 つ目の fold + Under-Sampling	648	3,942

また，データをミニバッチごとに分ける際にミニバッチ内のラベル比率を保つために Bagging を行う．Under-Sampling と Bagging の概要を図 3 に示す．

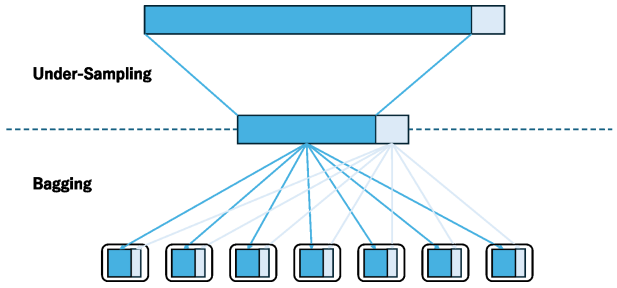

図 3 UnderSampling&Bagging の概要

(4) 新規の医療機関のデータに対する予測

新規の医療機関のデータに対する精度を検証するために特定の医療機関で得られたデータをテスト用データセットとして予測を行う．本稿で取り扱うデータセットは 7 つの医療機関から得られたものであるが，医療機関ごとに特徴や偏りがあるため，テスト用データセットの作成方法によって精度が大きく変化する．テスト用データセットの作成方法を変更した際の精度を図 4 に示す．

全ての医療機関から均等にデータを集め，テスト用データセットとした Cross Validation やデータセットの多くを占める福岡大学病院や九州大学病院のデータをテスト用データセットとして扱った場合と特定の医療機関のデータのみを使用した Test Dataset では精度に大きな差がある．これらの結果から本稿では特定の医療機関から得られたデータのみでテスト用データセットを作成し，精度を検証する．

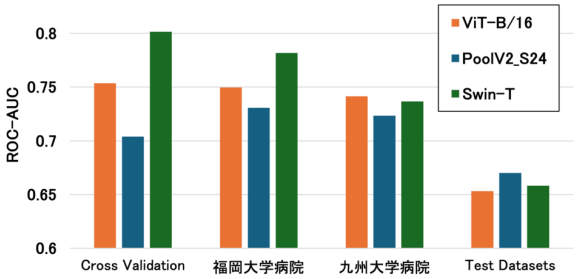

図 4 テスト用データセットに対する精度

5. 評価指標 ROC-AUC

モデルの性能や各手法を評価する指標として ROC-AUC (Area Under the ROC Curve) を用いる．ROC-AUC は 2 値分類タスクの評価指標として用いられる．ROC-AUC の値は 1.0 に近いほどモデルの予測精度が高いことを示し，0.5 に近いほど予測精度が低いことを示す．

6. Transformer をベースとしたモデル

本研究では，言語処理タスクにおいて多くの成功を収めた Transformer を画像処理分野に応用したモデルである VisionTransformer (ViT)，PoolFormer, SwinTransformer を用いる．

(1) VisionTransformer (ViT)

ViT は Transformer の Encoder を利用したモデルである [2]．画像をパッチに分けて Embedding し，Multi-Head Attention (MHA) を用いてパッチ間の関係を捉えている [3]．ViT の構造を図 5 に示す．

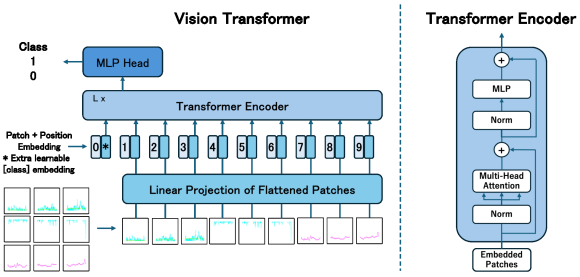

図 5 ViT の構造図

(2) PoolFormer

PoolFormer は階層的な構造を持ち，ViT の MHA を Pooling に置き換えたモデルである．Token mixer に当たる MHA を単純な Pooling に入れ替えることで MetaFormer (Patch Embedding, Token mixer, MLP) の構造が ViT に強く影響していることを示したモデルである [4][5]．PoolFormer の構造を図 6 に示す．

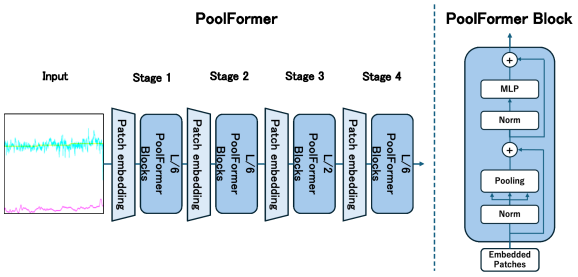

図 6 PoolFormer の構造図

(3) Swin Transformer

Swin Transformer は PoolFormer と同様に階層的な構造を持ち，ViT の MHA を Shifted Window Multi-Head Attention (W-MHA, SW-MHA) に置き換えたモデルである．Window の中で Attention を計算することで計算量を削減し，その後 Window を Shift させることで特徴を繋ぎ合わせている [6]．Swin Transformer の構造を図 7 に示す．

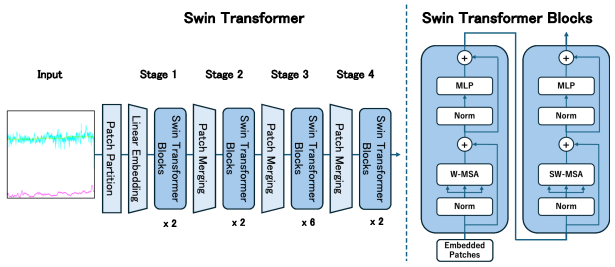

図 7 Swin Transformer の構造図

7. 提案手法

本研究では，Transformer をベースとしたモデルと CNN モデルを用いて以下の実験を行い，時系列データである CTG に対して有効な手法を探索する．すべての実験において seed 値を 3 回変更し，10-fold 交差検証 (Cross Validation: CV) を行うことでモデルの汎化性能を評価する．

(1) 信号処理による精度の検証

CTG を構成する FHR と UC に対して信号処理を行い，精度を検証する．FHR には測定器の装着不良が原因で生じたノイズや母体心拍の混入，欠損値など多くのアーチファクトが存在する [1]．また，一般的にノイズは高周波に含まれている．UC は腹部から陣痛計を装着し測定しているため，正確な値が記録されていない．産婦人科医は収縮の有無や間隔，UC に対する FHR の反応を評価している [1]．

a) FHR に対するローパスフィルタ

FHR に対してローパスフィルタを適用させ，精度を検証する．カットオフ周波数を段階的に設定することで高周波に含まれているノイズを除去する．FHR に対してローパスフィルタを適用させた CTG を図 8 に示す．

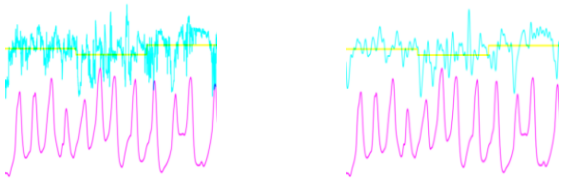

図 8-a 元画像 **図 8-b 適用後の画像**

図 8 ローパスフィルタ

b) UC に対する正規化

UC に対して正規化を行い，固定値倍することで精度を検証する．固定値を段階的に設定することで UC の大きさを変化させる．UC に対して正規化を行い，固定値倍した CTG を図 9 に示す．

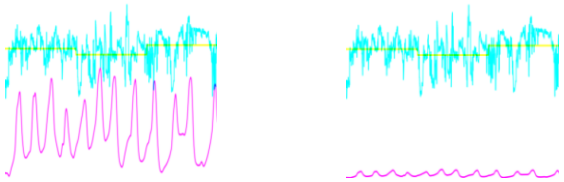

図 9-a 元画像 **図 9-b 適用後の画像**

図 9 正規化

(2) データ拡張による精度の検証

CTG に対してデータ拡張を行い，その有無で精度の検証を行う．また，本稿では時系列データに適したスリット型 CutOut とスリット型 CutMix を提案する．

a) CutOut と CutMix

CTG に対して Devrie[11] らが提案した CutOut と Yun[12] らが提案した CutMix を適用し，精度の検証を行う．CutOut は画像データの一部をランダムにマスクするデータ拡張である．CutOut を適用した CTG を図 10 に示す．以下の図では視認しやすくするために灰色の矩形でマスクしているが，学習時には白色の矩形でマスクする．

CutMix は CutOut で画像の一部をマスクする矩形を他の画像のパッチに置き換えたデータ拡張である．また，画像に対するパッチ面積の割合に比例してラベルを混合させる．CutMix を適用した CTG を図 11 に示す．

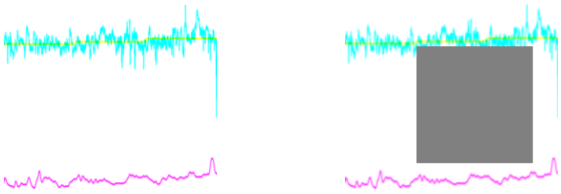

図 10-a 元画像　　図 10-b 適用後の画像

図 10 CutOut

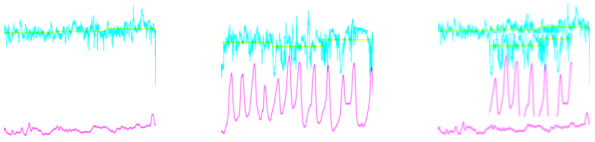

図 11-a 元画像　図 11-b 別の元画像　図 11-c 適用後の画像

図 11 CutMix

(3) スリット型 CutOut とスリット型 CutMix

CTG に対して新たに提案するスリット型 CutOut とスリット型 CutMix を適用し，精度の検証を行う．本稿で扱う CTG は時系列データであるため，同時刻の FHR と UC を同時にマスクすることが重要である．スリット型 CutOut はマスクする矩形の縦幅を画像の縦幅で固定したデータ拡張である．スリット型 CutOut を適用した CTG を図 12 に示す．また，CutMix についても同様にスリット型 CutMix を提案する．スリット型 CutMix を適用した CTG を図 13 に示す．

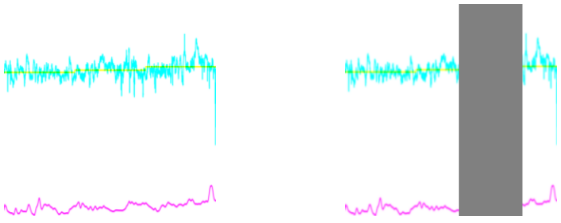

図 12-a 元画像　　図 12-b 適用後の画像

図 12 スリット型 CutOut

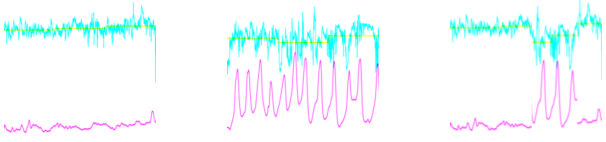

図 13-a 元画像　図 13-b 別の元画像　図 13-c 適用後の画像

図 13 スリット型 CutMix

(4) モデル構造の比較

実験を通して ViT, PoolFormer, SwinTransformer のパラメータ数が同程度であるモデル間で精度を比較し，CTG に対して有効なモデル構造を探索する．また，CNN モデルである ResNet[13][14] をベースラインモデルとして使用する．本実験で使用する各モデルのパラメータを表 5 に示す．

表 5 モデルのパラメータ数

Model Name	Parameter Count
ViT-S/16	22M
ViT-B/16	86M
PoolV2_S24	21M
PoolV2_M48	73M
Swin-T	28M
Swin-B	88M
ResNet50	25M
ResNet152	60M

8. 実験結果

(1) 信号処理の結果

FHR に対してローパスフィルタを適用した結果を図 14 に示す．図 14 に示すように，カットオフ周波数を 5 から 50 の間で変化させて高周波成分を取り除いても，精度に影響がないことがわかる．このことから，少なくとも学習したモデルにおいては，FHR の高周波成分には重要な特徴が含まれておらず，その結果，低周波帯域の波形のみを用いてハイリスクな出産を学習/予測するモデルを用いても，遜色のない精度が得られることがわかる．

また，カットオフ周波数が 0 の場合（この場合 FHR のグラフは平均値のみの直線である）に精度が大きく落ちている．そのため，用いたモデルにとって，ハイリスクな出産を予測するためには，FHR の大域的な形状は，重要な特徴であるといえる．

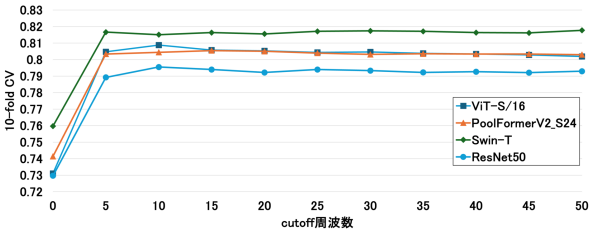

図 14 FHR に対してローパスフィルタを適用した結果

次に，UC に対して正規化を行い，固定値 (係数) 倍したものを入力データとして実験を行った時の予測精度の変化を図 15 に示す．固定値 (係数) を 10 から 120 の間で変化させたが精度に大きな変化は見られず，係数

が 0 の場合（この場合 UC のグラフは値が 0 の直線である）に精度が下がっている．以上の結果から，用いたモデルにとって，UC の絶対値自体は重要でない一方で，UC も考慮することが予測精度の向上に有効であることがわかる．

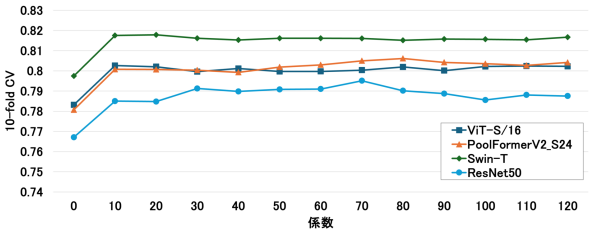

図 15 UC を正規化した結果

(2) データ拡張の適用結果

データ拡張を CTG に適用した結果を表 6 に示す．表 6 より，10-fold CV においては本研究で使用したデータ拡張はすべてのモデルで精度が向上した．また，CutMix は 6 つのモデルにおいて他のデータ拡張よりも高い精度を示した．テストデータセットに対してはすべてのモデルで精度が向上したデータ拡張はなかったが，スリット型 CutMix は 6 つのモデルにおいて他のデータ拡張よりも高い精度を示した．

(3) モデル間の精度比較

表 6 より，10-fold CV においては Swin-B が最も高い精度 (0.8310) を示した．一方で，テストデータセットに対しては ViT-B/16 が最も高い精度の (0.6848) を示した．また CNN モデルである ResNet50 はテストデータセットに対して 0.6758 と本稿で使用したモデルの中で 3 番目に高い精度を示した．

9. 結論

本稿文では Transformer をベースとした深層学習モデルを用いて CTG に対して有効な手法を探索した．FHR にローパスフィルタを適用した結果から FHR の高周波帯域には重要な特徴を含んでおらず，モデルは FHR の低周波帯域を特徴として扱い，ハイリスクな出産を予測していることが判明した．CTG に対してデータ拡張を行った結果，CutMix と新たに提案したスリット型 CutMix が CTG に対して有効な手法であり，高い精度を示した．特にスリット型 CutMix はテストデータセットに対して高い精度を示したことから，交差検定法では捉えられないような医療機関ごとのデータのバイアスに対しても比較的頑強であり，新規の医療機関のデータに対しても高い精度を期待できる．モデル間で精度を比較した結果としては，全体的に，Swin Transformer と VisionTransformer が CTG に対して高い精度を示した．そのため，本研究で用いたモデルの中ではこれらのモデル構造が CTG に対して有効であるといえる．

表 6 データ拡張を適用した結果

モデル+データ拡張	10-foldCV (mean ± std)	テスト
ViT-S/16	0.8024 ± 0.0247	0.6532
ViT-S/16 + CutOut	0.8080 ± 0.0248	0.6532
ViT-S/16 + スリット型 CutOut	0.8090 ± 0.0244	0.6499
ViT-S/16 + CutMix	**0.8149 ± 0.0251**	0.6544
ViT-S/16 + スリット型 CutMix	0.8143 ± 0.0212	**0.6646**
ViT-B/16	0.8093 ± 0.0246	0.6210
ViT-B/16 + CutOut	0.8168 ± 0.0267	0.6274
ViT-B/16 + スリット型 CutOut	0.8163 ± 0.0269	0.6509
ViT-B/16 + CutMix	**0.8240 ± 0.0213**	0.6434
ViT-B/16 + スリット型 CutMix	0.8228 ± 0.0211	**0.6848**
PoolV2_S24	0.7953 ± 0.0279	0.6701
PoolV2_S24 + CutOut	0.8085 ± 0.0269	0.6635
PoolV2_S24 + スリット型 CutOut	0.8096 ± 0.0252	0.6605
PoolV2_S24 + CutMix	**0.8136 ± 0.0260**	0.6496
PoolV2_S24 + スリット型 CutMix	0.8121 ± 0.0271	**0.6760**
PoolV2_M48	0.7481 ± 0.0591	0.6039
PoolV2_M48 + CutOut	0.7628 ± 0.0589	0.6056
PoolV2_M48 + スリット型 CutOut	0.7611 ± 0.0495	0.6346
PoolV2_M48 + CutMix	0.7665 ± 0.0398	0.5776
PoolV2_M48 + スリット型 CutMix	**0.7731 ± 0.0468**	**0.6423**
Swin-T	0.8164 ± 0.0265	0.6583
Swin-T + CutOut	0.8262 ± 0.0267	0.6469
Swin-T + スリット型 CutOut	0.8243 ± 0.0262	0.6601
Swin-T + CutMix	0.8287 ± 0.0274	**0.6616**
Swin-T + スリット型 CutMix	**0.8295 ± 0.0275**	0.6548
Swin-B	0.8193 ± 0.0258	0.6147
Swin-B + CutOut	0.8242 ± 0.0255	0.6487
Swin-B + スリット型 CutOut	0.8249 ± 0.0263	0.6466
Swin-B + CutMix	**0.8310 ± 0.0273**	**0.6576**
Swin-B + スリット型 CutMix	0.8300 ± 0.0256	0.6305
ResNet50	0.7878 ± 0.0277	0.6368
ResNet50 + CutOut	0.7963 ± 0.0233	0.6147
ResNet50 + スリット型 CutOut	0.7983 ± 0.0238	0.6722
ResNet50 + CutMix	**0.8205 ± 0.0231**	0.6708
ResNet50 + スリット型 CutMix	0.8193 ± 0.0252	**0.6758**
ResNet152	0.7778 ± 0.0248	0.6116
ResNet152 + CutOut	0.7873 ± 0.0283	0.6361
ResNet152 + スリット型 CutOut	0.7894 ± 0.0268	0.6068
ResNet152 + CutMix	**0.8229 ± 0.0241**	0.6518
ResNet152 + スリット型 CutMix	0.8227 ± 0.0269	**0.6562**

謝辞: 本研究を行うにあたり，終始ご熱心にご指導頂いた指導教員の柴田千尋准教授に厚く感謝を申し上げます．

参考文献

[1] 中井章人. 図説 CTG テキスト. 株式会社メジカルビュー社, 2016.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you

need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 2017.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[4] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10819–10829, June 2022.

[5] Weihao Yu, Chenyang Si, Pan Zhou, Mi Luo, Yichen Zhou, Jiashi Feng, Shuicheng Yan, and Xinchao Wang. Metaformer baselines for vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, No. 2, p. 896–912, February 2024.

[6] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, Vol. abs/2103.14030, , 2021.

[7] Zhidong Zhao, Yanjun Deng, Yang Zhang, Yefei Zhang, Xiaohong Zhang, and Lihuan Shao. Deepfhr: intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network. *BMC Medical Informatics and Decision Making*, Vol. 19, No. 1, pp. 286–, 2019.

[8] M. AnwarMa ' sum, P Riskyana Dewi Intan, Wisnu Jatmiko, Adila Alfa Krisnadhi, Noor Akhmad Setiawan, I Made Agus Dwi Suarjaya. Improving deep learning classifier for fetus hypoxia detection in cardiotocography signal. In *2019 International Workshop on Big Data and Information Security (IWBIS)*, pp. 51–56, 2019.

[9] Alkanan Mohannad, Chihiro Shibata, Kohei Miyata, Toshiro Imamura, Shingo Miyamoto, Hiroaki Fukunishi, and Hiroyuki Kameda. Predicting high risk birth from real large-scale cardiotocographic data using multi-input convolutional neural networks. *Nonlinear Theory and Its Applications, IEICE*, Vol. 12, No. 3, pp. 399–411, 2021.

[10] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114. PMLR, 09–15 Jun 2019.

[11] Terrance Devries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *CoRR*, Vol. abs/1708.04552, , 2017.

[12] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Young Joon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6022–6031, 2019.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, pp. 770–778. IEEE, June 2016.

[14] Ross Wightman, Hugo Touvron, and Herve Jegou. Resnet strikes back: An improved training procedure in timm. In *NeurIPS 2021 Workshop on ImageNet: Past, Present, and Future*, 2021.