Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

PU学習と相互情報量最大化を用いた染色体異常胚による過学習の抑制

長屋 雅士[1,a)]　澤田 祐季[2,b)]　佐藤 剛[2,c)]　澤田 富夫[3,d)]　杉浦 真弓[2,e)]　浮田 宗伯[1,f)]

概要：近年晩婚化のため，高齢出産をする女性が増えており，体外受精の需要が高まっている．体外受精の成功率向上のため，出産成功する可能性の高い良質な胚を選ぶ過程において，深層学習を用いる研究が進められている．従来の研究では，出産成功胚を正例（Positive），出産失敗胚を負例（Negative）とラベル付けして，PN 分類するものがほとんどである．しかし，高齢出産に伴い増加する胚の染色体異常が起きている場合，見た目が出産成功胚とほぼ同じであっても出産失敗に至ってしまう．つまり，出産失敗胚には見た目の質が良いものと悪いものが共存してしまっている．そのため，出産失敗胚すべてを負例として PN 学習してしまうと，分類に悪影響を及ぼす可能性が高い．この悪影響を抑制するため，出産失敗胚に負例ラベルを与えるのではなく，出産成功と失敗のどちらの可能性もある「ラベルなし（Unlabeled）」とした PU 学習の利用を提案する．またこの PU 学習をベースにし，ランク学習への拡張による PU 学習の改善，相互情報量最大化による距離学習，時間方向への滑らかさ制約によるタイムラプス画像の判断根拠の出力安定化についても行った．本稿では，以上で述べた手法が出産成否の分類問題において有用であることを実験的に示した．

キーワード：体外受精，染色体異常，Positive-Unlabeled 学習，ランキング学習，相互情報量最大化，判断根拠の可視化

1. 序論

近年晩婚化のために高齢出産をする女性が増えており，体外受精の需要が高まっている．現在の体外受精においては，複数の胚を培養し，それらの形状や細胞の均一性を専門医が総合的に判断し，最も質の良い胚を体内に移植する．この胚の見た目が，出産成否をわける最も大きな特徴であるにも関わらず，その見た目ですら出産成功する胚と失敗に至る胚の見た目が似ていることから，出産成功に至る質の良い胚を確実に選択することは，専門医にとっても非常に難しい問題である．そこで画像分類のタスクにおいて高い性能をもつ深層学習を，良質な胚の分類に応用する研究が進められている [1], [2].

Khosravi らは Veeck と Zaninovic の胚形状評価枠組み [3] に従い，胚鑑定士がすべての胚にラベル付けを行うことで，良好胚と不良胚の 2 クラス分類を行った [1]. しかし 1 つ 1 つの胚に人手でラベル付けを行っていることから，莫大なアノテーションコストに加え，胚鑑定士同士でラベルのばらつきが生じてしまう．また，胚鑑定士が判断した見た目によってラベルがつけられていることから，あくまで胚鑑定士の予測を再現するだけとなってしまう，一方宮城らは人手ではなく，実際に得られた出産成否結果に基づいて，出産成功胚を正例（Positive）出産失敗胚を負例（Negative）とラベル付けを行い，PN 分類を行った [2]. これにより，出産成功しうる胚の特徴そのものを学習できる可能性がある．しかしこのラベル付けの問題点として，もし仮に出産成功と同じ見た目の胚であったとしても，染色体異常が含まれていた場合，出産成功に至る可能性は低い．つまり失敗とラベル付けされた胚は，見た目の質が良い胚と悪い胚が共存してしまっている．以上のことから従来手法では，ラベル情報が不正確であり，分類に悪影響を及ぼす可能性が高い．そこで我々はこの悪影響を抑制するため，出産失敗胚に負例ラベルを与えるのではなく，出産成功と失敗のどちらの可能性もある「ラベルなし（Unlabeled）」とした PU 学習の利用を提案する．そしてこの PU 学習をベースにさらに 3 つの手法を加えることでさらなる分類精度向上

[1] 豊田工業大学
[2] 名古屋市立大学
[3] さわだウイメンズクリニック
[a)] sd19437@toyota-ti.ac.jp
[b)] yuukidec27@yahoo.co.jp
[c)] og.sato@med.nagoya-cu.ac.jp
[d)] tomysawada@ybb.ne.jp
[e)] og.sato@med.nagoya-cu.ac.jp
[f)] ukita@toyota-ti.ac.jp

を目指す．本論文の新規性は以下である．

(1) PU 学習による誤ったラベリングへの対応：
本来 PU 学習は，正例とラベルなしデータのみ存在する半教師あり学習という問題設定で用いられる．しかし提案手法では，負例ラベルの信頼性の低さによる悪影響を抑制するために，あえてすべての出産失敗胚をラベルなしとして扱い学習を行う．

(2) AUC 最適化による PU 学習
PU 損失による深層学習モデルの最適化は，ハイパーパラメータに依存する．そこでランキング学習に用いられる AUC 損失の最適化も行うことで PU 学習を改善した．

(3) 相互情報量最大化による距離学習
出産成否の予測問題において，正例と負例の各クラス間で見え方の差が小さく，適切な分離境界の決定が困難であることが想定される．そこで相互情報量最大化を行うことでクラス内分散を小さく，クラス間分散を大きくし，より適切な分離境界が決定されることを期待する．

(4) 出産成否分類問題における判断根拠の可視化
医療分野においては特に，深層学習モデルの分類根拠を明らかにすることが重要である．提案手法では判断根拠を可視化しつつ，それを分類にも利用する Attention Branch Network（ABN）[4] をベースに時間方向の滑らかさ制約を加える．これにより，今回のようなデータ数に限りがあるような問題設定であっても，安定したタイムラプス画像の可視化を実現した．

2. 関連研究

2.1 PU 学習

近年の深層学習技術の発展により，医療分野においてもその応用が広がっている．医療分野においてはデータ収集の難しさだけでなく，データに正解情報を付与する際に高度な専門知識を要することから，深層学習に必要十分な量のデータが存在するケースは少ない．そこでラベルありデータを基に，ラベルなしデータも有効に利用した半教師あり学習が行われることが多く，例えば胸部 X 線セグメンテーション [5] や多発性硬化症の検出 [6] に用いられている．この半教師あり学習のひとつに PU 学習がある．PU 学習とはラベル付けされた正例（Positive）と，正例と負例のどちらの可能性もあるラベルなし（Unlabeled）が存在するという設定で分類を行う．この PU 学習はラベルなしデータをどのように扱うかで 2 つの手法に分けられる．1 つはラベルなしデータ中に存在する負例を特定し，そのデータにラベル付けを行った後に，通常の PN 学習を行う手法である [7], [8]．2 つめはラベルなしデータを小さな重みのかかった負例データとして扱う手法である [9], [10]．この文献 [9], [10] によって，負例が存在しない PU 学習でも，PN 学習の性能を超えることが実験的にも理論的にも証明された．さらに [10] では，従来線形なモデルのみで利用されていた PU 学習を，深層学習器を用いた非線形なモデルにも拡張した．

2.2 ランキング学習

深層学習モデルを用いた PU 学習のひとつに，非負 PU リスクを用いた手法がある [10]，しかしこの手法では，いくつかのハイパーパラメータ調整が必要なことから，その性能が人手に依存してしまうという問題点があった．一方ランキング問題で用いられる評価指標である AUC を最適化することで，PU 分類を行うアプローチもある [11]．しかし AUC は微分不可能であることから，深層学習モデルをそのまま最適化することはできない．そこで微分可能な形に近似した AUC 損失が提案されている [12], [13]，これらはそれぞれ，ウィルコクソンの順位和検定，多項式近似を用いて微分可能な形に近似している．しかしこれらの近似による AUC 損失では，性能が不十分であったり，計算コストが大きいという問題点があったが，文献 [14] では上界下界を導入することによって，効率のよい近似を実現した．

2.3 PU 学習における距離学習

出産成否の分類問題において，出産成功と出産失敗では見え方の差がほとんどなく，クラス間の分散がかなり小さいことが想定される．このような状況では分離境界を適切に決定することができず，分類性能が下がる懸念がある．そこでクラス内分散を小さくし，クラス間分散を大きくすることを目指した距離学習が提案されている [15]．この手法ではまず基準となるデータを用意し，その基準データと同じクラスのサンプルと異なるクラスのサンプルでそれぞれペアを形成する，その後特徴量空間において各ペア内のサンプル同士で距離を計算し，同一クラスペアの距離は小さく，逆に異なるクラスペアの距離を大きくするように学習を行う．これによりクラス内分散を小さく，クラス間分散を大きくすることができる．しかしこの手法のようにペアを作るには正しいラベル情報が必要となるため，今回のようなラベルなしデータが存在する問題設定では用いることができない．

一方異常検知のための PU 学習において，距離学習を導入した手法が提案された [16]．この手法ではまずターゲット画像を 1 つ決め特徴量空間上で，ラベルなしデータ間と正例間それぞれとの距離を計算する．その距離計算結果を基にハッシングフィルタリングを行い，ラベルなしデータにラベル付けをすることで [15] の距離学習を可能とした．この手法は異常検知のように，ラベルなしデータの多くが，正例ではない問題設定においては有効である．しかし出産成否問題においては，ラベルなしデータに多くの正例と同

Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

じ見え方の画像が含まれている可能性があり，有効な距離学習は難しい．また距離計算を行い，フィルタリング処理も行うため，追加の処理が多く計算コストが大きいという問題もある．そこで，距離計算とラベル情報が不要な相互情報量最大化による距離学習を提案する，

2.4 判断根拠の可視化

深層学習モデルの判断根拠を解釈する研究が進められている [4], [17], [18]．これらの可視化手法は，勾配計算を用いたボトムアップ型と，出力値を利用するトップダウン型のおおきく 2 つに分けられる．ボトムアップ型の例として Gradient-weighted Class Activation Mapping(Grad-CAM) が挙げられる [18]．Grad-CAM では，ある特定クラスの誤差を逆伝搬し，その予測結果に寄与した特徴量マップの重みを算出，元の特徴量マップとの積和演算により注視領域の可視化画像を出力する．一方トップダウン型のひとつである CAM では，予測クラス数分の特徴量マップに対し，Global Average Pooling を施すことで，各特徴量マップがどれだけ予測に寄与したかを表す重みを算出することができる [17]．しかし CAM は可視化のためにネットワーク構造を大きく変更する必要があり，タスクによっては性能低下を招く可能性があった．この問題を解消するために ABN[4] では，CAM をベースとした Attention 機構を導入し，出力されたアテンションマップを予測にも利用することで，予測根拠を可視化しつつ，分類性能を向上させることに成功した．

こうした手法を医療画像にも適用し，判断根拠を明確にすることで，患者に安心して治療を受けてもらえるため，医療画像においてもその研究が進められている．例えば癌の分類問題 [19] であったり，MRI 画像からパーキンソン病を予測する問題 [20] などが挙げられる．しかしこれらの従来研究では，単体の画像に対してのみ有効であった．したがって，今回のようなタイムラプス画像に対応した手法を提案する必要がある．

3. 提案手法

3.1 PU 学習

一般的な PN 学習であれば，出産成功胚を正例，出産失敗胚を負例とラベル付けを行う．しかし前述のように，仮に見た目の質が良い胚であっても染色体異常が含まれてしまっていると出産成功に至らない．すなわち出産失敗胚には見た目の質が良いものと悪いものが共存してしまっている．このような状況下で一般的な PN 学習を行ってしまうと，**図 1** 左の PN 学習に示すように，正例サンプルが多く含まれる分布中の，出産失敗とラベル付けされたサンプルもうまく分離しようとするため，分離境界が正例分布に入り組んでしまう．すなわち過学習が引き起こされ，本来出産成功する可能性のある見え方の良い胚に対しても，出産失敗と予測してしまう恐れがある．この本来出産成功するはずの胚を見落とすことは，胎児を獲得する機会を失い，時間的な損失だけでなく，再治療にかかる身体的，金銭的負担が大きくなる．したがって出産成否の予測において，見た目の質の良い受精卵を漏れなく，確実に予測することが非常に重要である．

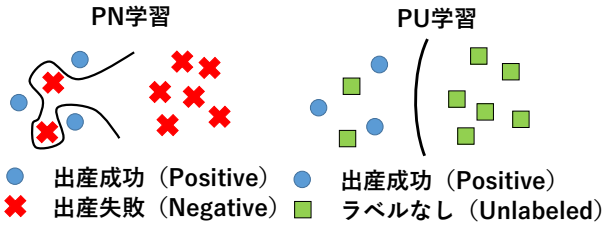

図 1 PN 学習（左）と PU 学習（右）を行ったときの分離平面のひかれた方の違い．PN 学習では出産成功と見た目が同じ染色体異常による出産失敗画像によって，分離平面が正例分布に入り組んでしまい過学習を引き起こしてしまう．一方 PU 学習を行い，出産失敗データすべてラベルなしとすることで過学習を抑制でき，見た目の良しあしに基づいて分離平面を引くことができる．

そこで我々は PU 学習を用いることを提案する．本来 PU 学習とは，正例のみが利用でき，ラベルが不明なラベルなしデータも存在するという問題設定において用いられる．しかし我々は出産失敗した胚に対して，負例とラベル付けするのでなく，あえてそれらをすべて Unlabeled として PU 学習を行う．これにより，見た目の質は良いが，染色体異常によって出産失敗に至った胚に対し，出産失敗と学習することがなくなり，見た目の質の良さに基づいて分離境界を引くことができるようになる（図 1 右の PU 学習）．こうして見た目の質が良い胚については，出産成功と予測できるように学習が進むため，胎児を獲得する機会を増やすことができる．以上のことから PU 学習は，深層学習モデルの過学習を防ぐという観点に加え，胎児獲得の機会を逃さないという 2 つの観点から有用である．

3.2 ランキング学習を拡張した PU 学習

本節では，ランキング学習を拡張し，文献 [10] で提案された非負 PU 学習を改良する．まず非負 PU 学習においては以下のリスクを最小化する，

$$\tilde{R}_{pu} = \pi_p \hat{R}_p^+(g) + max\{0, \hat{R}_u^-(g) - \pi_p \hat{R}_p^-(g)\} \quad (1)$$

ただし

$\hat{R}_p^+(g) = (1/n_p) \sum_{i=1}^{n_p} l(g(x_i^p), +1)$
$\hat{R}_p^-(g) = (1/n_p) \sum_{i=1}^{n_p} l(g(x_i^p), -1)$
$\hat{R}_u^-(g) = (1/n_u) \sum_{i=1}^{n_u} l(g(x_i^u), -1)$

であり，それぞれミニバッチ内において，どれだけ正例サンプルを正例と予測できたかを評価する損失の平均，どれだけ正例サンプルを負例と予測したかを評価する損失の平均，そしてどれだけラベルなしデータを負例と予測したか

Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

を評価する損失の平均を表す．また π_p は正例サンプルの事前分布を表す．ここで式 (1) 第 2 項目 $\hat{R}_u^-(g) - \hat{R}_p^-(g)$ の学習が進み，負の値になるとラベルなしデータをすべて負例と予測するようになってしまい過学習に陥る．それを避けるため，第 2 項目が 0 より小さくならないように制約をかけて学習を行う．これにより深層学習モデルにおいても過学習に陥らない PU 学習を可能とした．

上述の非負 PU 学習の枠組みに，AUC にて性能評価されるランキング学習を導入する．この AUC を直接最適化することで，分類問題で一般的に用いられるクロスエントロピーを用いた場合よりも，分類性能が向上したことや [14]，PU 学習の分類性能自体が向上したことも示されている [21]．以上より，非負 PU 学習と同時に AUC を最適化することで，今回の分類が難しい出産成否問題に対して対応できることが期待できる．

今回はランキング学習を評価する指標の中で，AUC-PR(Precision-Recall) を採用する．この AUC-PR は正例と負例のサンプルが不均衡な場合であっても，上手く性能評価できる指標であるためである．今回の出産成否問題においては，出産成功した数が出産失敗した例よりも明らかに少ないため，学習させる際に正例と負例で不均衡になってしまうことから，AUC-PR が適している，文献 [14] では，この AUC-PR をよりうまく表現した損失関数が提案されており，様々なネットワークに利用可能である．

PR カーブは分類器からスコアが出力された後，分類結果を分けるスコアの閾値を変化させることで描かれる．そして AUC-PR は recall を固定させたときの precision の積分値として表現される．つまり，recall=α と固定し precision を計算した後，α を変化させたときの precision の合計値が AUC-PR となる．ここで α が一定間隔 k で変化すると仮定し，$\alpha_t = \frac{k}{t}$ とする．こうして [14] に従い，以下の代理損失を AUC-PR の最適化に利用する．

$$\underset{\lambda_1 \ldots \lambda_t}{max} \sum_{t=0}^{k} \Delta_t((1+\lambda_t)\mathscr{L}_+(f,b_t) + \lambda_t \frac{\alpha_t}{1-\alpha_t}\mathscr{L}_-(f,b_t) - \lambda_t|Y_+|) \quad (2)$$

λ_t と b_t はそれぞれラグランジュの未定乗数と t 番目の閾値を表す．$\mathscr{L}_+(f,b_t)$ と $\mathscr{L}_-(f,b_t)$ はそれぞれ正例と負例に対しどれだけ誤った分類をしたかを表す．この代理損失を最小化することで AUC-PR の最適化を行う．

3.3 相互情報量最大化による距離学習

2.3 節で述べたように，出産成否問題を PU 学習するにあたり，有効な距離学習はまだ提案されていない．そこで相互情報量最大化を用いた距離学習を提案する．文献 [22] では，教師なしクラスタリングのためにまず，ある画像とそれに回転などの変換を加えた 2 枚のペアで，以下の相互情報量を計算する．

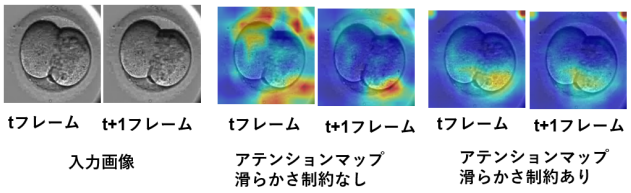

図 2 時間方向に滑らかさ制約をかけてない場合，時刻 t と $t+1$ で見え方がほとんど変わらないにもかかわらず，注視領域が大きく異なる（図中央）．一方滑らかさ制約を加えることで注視領域はほとんど一致しており，安定した出力が得られた．（図右）

$$I(z,z') = H(z) - H(z|z') \quad (3)$$

式 (3) 右辺第 1 項目が大きくなるほどデータ分布の広がりが大きくなり，逆に第 2 項目が小さくなるほど同じクラスである，似た見え方のデータ分布の広がりが小さくなる．すなわち相互情報量を最大化することは，クラス間分散を大きくし，クラス内分布を小さくすることと解釈できる．加えて式 (3) の計算には，ある画像とそれに変換を加えた画像のみが必要であり，正解データを必要としない，すなわち，相互情報量最大化はラベル情報がなくとも距離学習の効果を期待することができ，ラベルのない PU 学習を行う今回の手法に適している，今回は文献 [22] で提案された，式 (3) の IIC 損失を用いることで相互情報量を最大化する．

3.4 出産成否分類問題における判断根拠の可視化

アテンションマップの出力を安定化させるため，時間方向に滑らかさ制約を加えた学習を行う．今回の問題においては，胚の細胞分裂によって，見た目が大きく変化するのは約 1 日ごとであるため，撮影間隔である 10 分から 15 分程度の時間経過では，見た目が大きく変化することはほとんどない．実際**図 2** 左のように，時刻 t と $t+1$ で見え方がほとんど変わらないことが確認できる．しかし画像単体での出力を見ると，モデルの注視領域を示すアテンションマップは大きく異なってしまっている（図 2 中央）．これらの不安定な出力は，さらに大量のデータがあれば解決されることが期待されるが，データの集まりにくさから，その数には限りがある．そこで時刻 t と $t+1$ の出力結果に対して，平均二乗誤差を取り最小化することで，時間方向に対し滑らかな出力になるような制約をかけて学習を行う．これにより時刻 t と $t+1$ で注視領域がほとんど等しくなり，安定した出力結果を得ることができる（図 2 右図）．

3.5 実装詳細

図 3 に今回用いたネットワークを示す．これは ResNet56 をベースとした Attenton Branch Network [4] に対し，ImageNet[23] にて事前学習を行ったモデルである．ResNet56 は第 45 層で 2 分割されており，前段は特徴量抽出器とし

Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

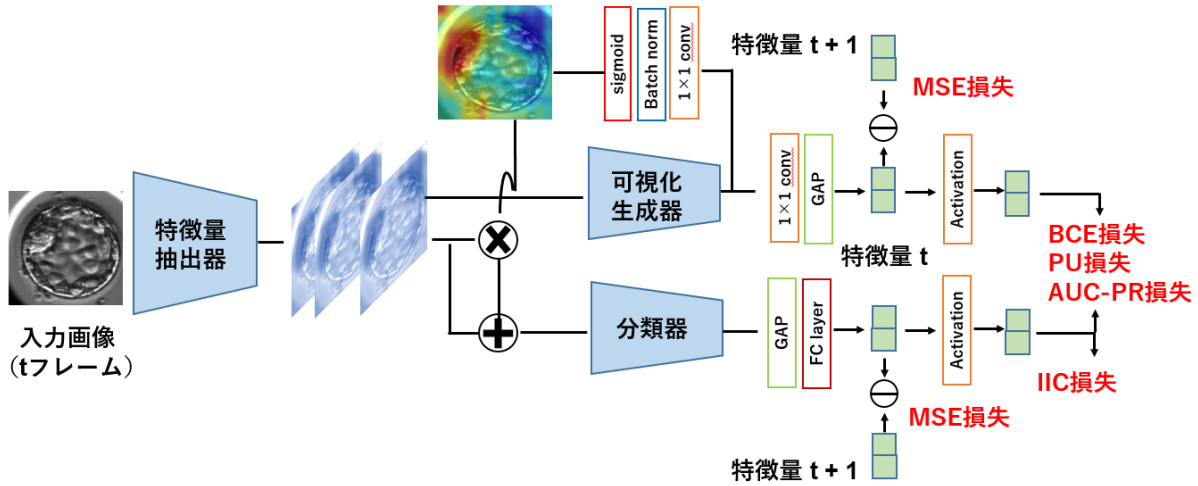

図 3 今回用いたネットワーク図．[4] のネットワークをベースとし BCE 損失，非負 PU リスク，AUC-PR 損失，IIC 損失，MSE 損失の最適化を行う．

表 1 データセットの分割詳細

データ	1	2	3	4	5	計
出産成功数	28	28	28	28	28	140
出産失敗数	100	100	101	101	101	503

て，後段は分類器として用いられる．また，可視化生成器は 10 層の畳み込み層で構成される．

可視化生成器と分類器からの出力それぞれに対し損失を計算し最適化を行う．用いる損失関数は Binary Cross Entropy（BCE），式 (1) の非負 PU リスク，式 (2) の AUC-PR 損失，式 (3) の IIC 損失を計算し最適化を行う．またそれぞれの損失計算の直前で異なる活性化関数を利用しており，BCE と IIC 損失，非負リスクに用いたのはそれぞれ softmax 関数，sigmoid 関数である．AUC-PR 損失に対しては活性化関数は施さない．さらにアテンションマップの出力安定化のために，時刻 t と $t+1$ のスコアで平均二乗誤差をとる．

4. 実験

4.1 データセット

本実験では．名古屋市立大学とさわだウイメンズクリニックにて治療が行われた胚のタイムラプス画像を用いる．名古屋市立大学において Primo Vision[18] にて 10 分間隔で 101 シーケンス（出産成功：8，出産失敗：93），さわだウイメンズクリニックにおいて Embryoscope[19] にて 15 分間隔で 542 シーケンス（出産成功：132，出産失敗：410）撮影された．（**図 4**）

また画像サイズはそれぞれ 250×250 ピクセルと，500×600 ピクセルであった．この撮影タイムラプス画像に対し，胚の位置を不変とするため，人手で画像中央に胚が来るようにクロップ処理を行う（**図 5**）．その後リサイズ処理を施し，224×224 の画像に変換する．こうして成形した計 643

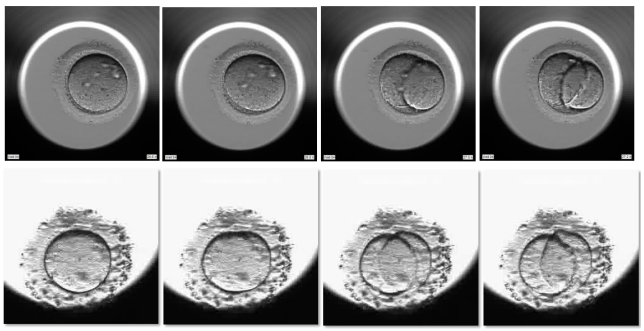

図 4 撮影されたタイムラプス画像．
（上段：さわだウイメンズクリニック，下段：名古屋市立大学）

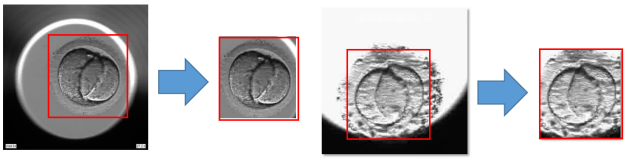

図 5 実際に行ったクロップ処理．これにより，どの画像においても中央に胚が位置するようにする．

シーケンスのタイムラプス画像を，出産成功と失敗の比率が等しくなるように 5 分割し交差検証を行う（表 1）．まず 5 分割されたうちの 1 つを検証データ，残りを学習データとし，ハイパーパラメータ（学習率，エポック数等）を決定する，次に分割された残りの 4 つのうち 1 つをテストデータ，残りを学習データとし，上で決めたハイパーパラメータを基に学習する．これを 4 つのテストデータすべてに対し行い，最終結果の平均値を最終的な性能評価に利用する．また今回のテストは，動画をフレーム分割した画像 1 枚 1 枚全てを用いて評価を行う．

Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

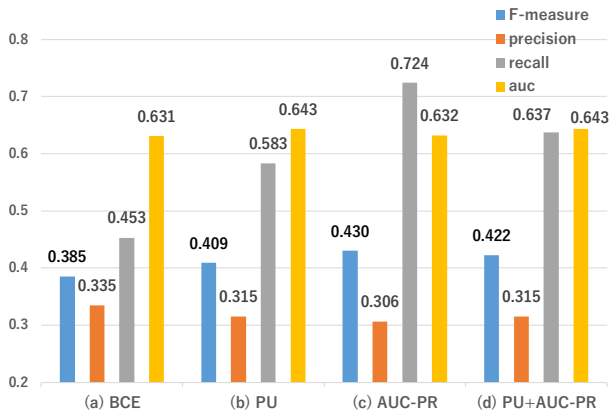

図 6　BCE, PU リスク, AUC-PR, PU Risk+AUC-PR をそれぞれ最適化した際の F 値 (青), precision(橙), recall(灰), AUC(黄) をそれぞれグラフ化した.

4.2　実験結果

各手法の性能評価として，正例をどれだけもれなく見つけることができたかを表す Recall，どれだけミスなく正例と予測できたかを表す Precision，Recall と Precision の調和平均をとった F 値，そして閾値に依存せずモデルの性能評価が可能な AUC の 4 つの指標を用いる．また Recall，Precision，F 値を計算する際に必要な，出産成否を分ける閾値については検証データによって決定する．

ベースラインである BCE，提案手法である非負 PU リスク，AUC-PR，非負 PU リスク+AUC-PR それぞれを最適化したテスト結果を**図 6** に示す．まず図 6(a) と (b) を比較すると，PU 学習を行うことで F 値が 0.024 向上した．その中で precision は 0.02 低下したものの，recall について 0.13 と大幅に向上した．これは図 1 に示した通り，染色体異常胚の悪影響を抑え，見た目のよしあしに基づいて分離境界を引くことができたと推測できる．つまり見た目の質が良いものはなるべく漏れなく正例と予測するように学習が進んだことにより，recall が上昇したと考えられる．逆に染色体異常胚のような，出産失敗だが見た目の質が良い胚についても同様に正例と予測してしまうことから precision が低下した．さらに AUC-PR を最適化した場合にも，前述と同様のことがいえる．しかし (c) は (b) と比べて，ハイパーパラメータが少ないことから，よりうまく最適化することができ，F 値や recall を大幅に上昇させることができたと考えられる．(d) においては PU 学習と AUC 最適化を同時に行うことで，PU リスク単体を最適化したときに比べ，precision を維持しつつ，recall を 0.054 向上させることができ，F 値も 0.013 向上した．したがって PU リスクと AUC-PR を同時に最適化することで，PU 学習の性能向上を確認することができた，最後に閾値によらないモデルの性能を評価する AUC についても，PN 学習の (a) よりその他の手法のほうが高い値になったことからも，提案手法の有効性を示すことができた．

表 2　IIC 損失の有無による比較．IIC の有無で比較した際，高い値の方の数字に対して赤字で強調した．

最適化損失関数	AUC	precision	recall	F 値
BCE	0.631	0.335	0.453	0.385
BCE+IIC	0.628	0.317	0.518	0.393
PU	0.643	0.315	0.583	0.409
PU+IIC	0.642	0.307	0.613	0.409
AUC-PR	0.632	0.306	0.724	0.430
AUC-PR+IIC	0.632	0.303	0.741	0.430
PU+AUC-PR	0.643	0.315	0.637	0.422
PU+AUC-PR+IIC	0.638	0.302	0.674	0.417

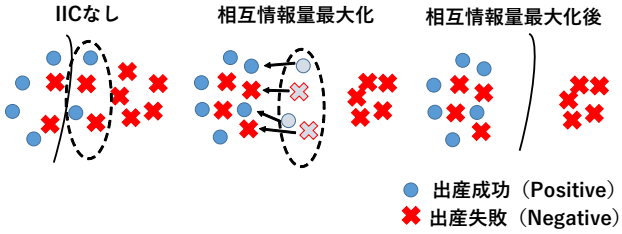

図 7　相互情報量最大化によるイメージ図．相互情報量を最大化することで，見た目が近いもの同士でクラスタを形成する効果が働く．すなわち，明らかに見た目の質が良くない胚画像で分布を形成し，さらに正例分布と負例分布の中間に位置するような胚画像については，正例分布にまとまると想定される．

続いて相互情報量最大化による影響を検証するために，IIC 損失の有無による 4 指標の値の変化を表 2 にまとめる．BCE，PU リスク，AUC-PR，PU リスク+ AUC-PR どの損失関数を最適化した場合でも，IIC 損失を追加することで AUC や F 値が大きく変化することはなかった．しかし precision がどの手法においても 0.01 から 0.02 と僅かに低下しているものの，recall については 0.02 から 0.06 程度上昇していることが確認できる，これは相互情報量最大化による，見た目が近い画像同士でクラスタを形成しようとする効果によるものと考えられる．その概念図を**図 7** に示す．もともと図 7 左のように，正例と負例が混ざり合っており，正例を逃すような位置に，分離境界が引かれてしまうことが想定される．しかし相互情報量も最大化することで，見た目の質が明らかに悪い胚画像で一つの分布を形成し，正例分布と負例分布の中央に位置するような胚画像については，ある程度見た目の質が良いことから，正例分布の方にまとまるように学習が進む．その結果，図 7 右図のように分離境界がひかれ，正例に近い微妙な負例に対する誤分類は増えてしまうものの，より多くの正例を見つけることができる．以上のことから precision は僅かに低下するものの，recall については上昇したものと推測できる．

まとめると提案手法によって，モデルの性能を向上させることができ，さらに出産成功する可能性の高い胚を漏れなく予測するような学習ができていることを実験的に示せた．3.1 節でも述べたように，出産成否問題においては，胎児獲得の機会を逃さないことが最も重要であることから，

Vol.2020-CG-180 No.13
Vol.2020-DCC-26 No.13
Vol.2020-CVIM-223 No.13
2020/11/5

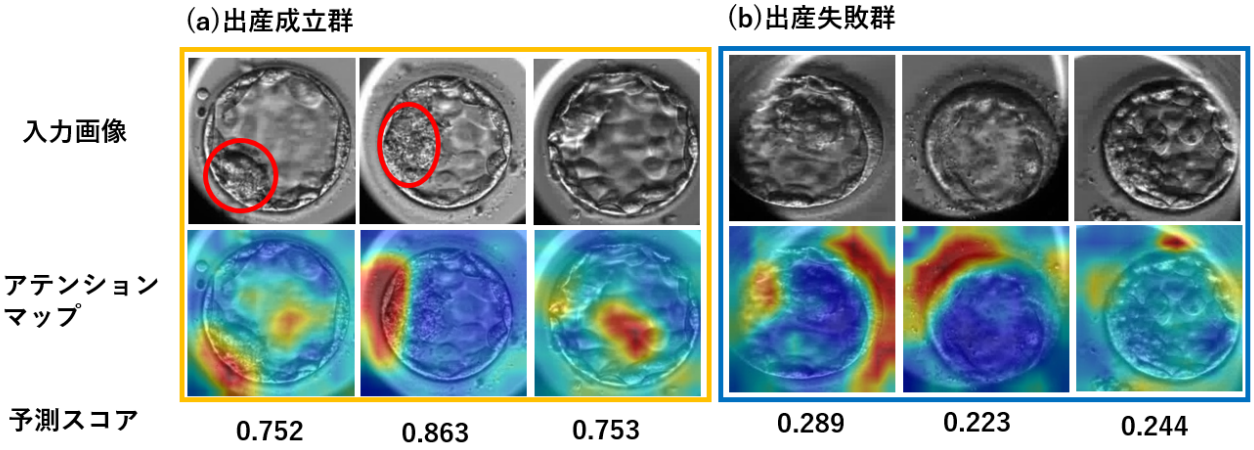

図 8　正例 (a) と負例 (b) の画像を ABN に入力した際のアテンションマップの例．どの画像においても受精後 115h 経過した際の画像であり，赤く発火した領域に特に注視している．

PU 学習そして，相互情報量最大化は有用であると結論付けられる．

最後に ABN による可視化画像を図 8 に示す．図 8 の (a) と (b) はそれぞれ実際に出産成功した胚を正しく正例と予測できた例と，出産失敗した胚を正しく負例と予測した例である．また画像はすべて受精後 115 時間経過時に撮影されたものである．まず (a) において，左上の図の赤枠で囲まれた部分のような，小さな塊に注視しているようなアテンションマップが得られた．この赤枠部分は内細胞塊と呼ばれ，Veeck と Zaninovic の胚形状評価枠組み [3] の評価対象のひとつである．こちらの内細胞塊の密度が大きいほど胚の評価としては高く，出産成功する可能性が高い．この領域の密度が大きい (a) 群左や中央のような画像において，スコアは 0.752，0.863 と高い値で出産成功予測と予測しており，おおよそ専門医の知見と一致している．しかし右の画像のように，内細胞塊ではない領域に注目している画像もあり，すべての画像において内細胞塊に注視しているというわけではないこともわかる．以上より内細胞塊が出産成功予測に共通した特徴であると結論付けることはできなかった．また (b) では，胚を囲うような領域に注視している画像が見られた．このような領域には，胚形状評価枠組みのひとつである栄養外胚葉が位置しており，この密度が高いほど胚の評価は高く出産成功数する可能性が高い．しかしアテンションマップを見ると，注視している栄養外胚葉の密度は小さく，実際にスコアは 0.3 以下とかなり低く予測していることから，こちらも専門医の知見と一致する．しかし右図のように胚全体をまんべんなく注視している画像も存在しており，一概に栄養外胚葉のみに注目しているというわけではなかった．以上まとめると，専門医の知見と一致するような点に注視している画像もあったが，そうでない領域に注目している画像もあり，出産成否を分ける共通した特徴をアテンションマップから読み取ることはできなかった．今後 ABN に学習させるデータ数を増やすことで，より共通した特徴をみつけられるようになる可能性はある．

5. 結論

今回は出産成否の予測問題において PU 学習を採用することで，予測性能を向上させることに成功した．さらに文献 [14] の AUC-PR も同時に最適化することで，PU 学習をさらに改善することができた．また IIC 損失によって，ラベルがないような問題設定においても，距離学習の効果が働き，質の良い胚を逃さないように学習が進んだことが確認できた．よって出産成否問題において，PU 学習，AUC-PR 最適化，相互情報量最大化は有効であるといえる．

今後は廃棄胚も有効活用した PNU 学習に取り組む．今回の問題では，出産成功と出産失敗の 2 クラス分類を行ったが，そのほかに廃棄された胚も存在する．この廃棄胚は見た目の質が悪く，移植するに値しないため廃棄された胚のことである．しかしこの廃棄胚は出産失敗の特徴がより現れた胚であることから，深層学習モデルの表現獲得に大きく役立つと考えられる．加えて特徴量空間上で出産失敗胚に対し，この出産成功胚と廃棄胚それぞれの距離を計算し比較することで，出産失敗胚が単純に見え方の質が悪かったのか，それとも見え方の質はよいが染色体異常によって出産失敗に至ったのかを特定することができる．したがって PNU 学習の枠組みに，より効果的な距離学習を導入することができる．

また今回は胚の形態に基づく評価指標に基づいて，CNN をベースとした手法を提案したが，胚の形態だけでなく，二前核の出現や細胞分裂のタイミングが出産成否に関わるという報告もある [24]．そこで Recurrent Neural Network などの時系列モデルを利用し，時間に関する特徴も考慮することで，より出産成否を正確に認識できる可能性もある．

参考文献

[1] Pegah Khosravi, Ehsan Kazemi, Qiansheng Zhan, Jonas E Malmsten, Marco Toschi, Pantelis Zisimopoulos, Alexandros Sigaras, Stuart Lavery, Lee AD Cooper, Cristina Hickman, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ digital medicine*, 2(1):1–9, 2019.

[2] Yasunari Miyagi, Toshihiro Habara, Rei Hirata, and Nobuyoshi Hayashi. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reproductive medicine and biology*, 18(2):204–211, 2019.

[3] Lucinda L Veeck and Nikica Zaninovic. *An atlas of human blastocysts*. CRC Press, 2003.

[4] Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi. Attention branch network: Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10705–10714, 2019.

[5] Gerda Bortsova, Florian Dubost, Laurens Hogeweg, Ioannis Katramados, and Marleen de Bruijne. Semi-supervised medical image segmentation via learning consistency under transformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 810–818. Springer, 2019.

[6] Christoph Baur, Shadi Albarqouni, and Nassir Navab. Semi-supervised deep learning for fully convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 311–319. Springer, 2017.

[7] Xiaoli Li and Bing Liu. Learning to classify texts using positive and unlabeled data. In *IJCAI*, volume 3, pages 587–592, 2003.

[8] Bing Liu, Wee Sun Lee, Philip S Yu, and Xiaoli Li. Partially supervised classification of text documents. In *ICML*, volume 2, pages 387–394, 2002.

[9] Marthinus C Du Plessis, Gang Niu, and Masashi Sugiyama. Analysis of learning from positive and unlabeled data. In *Advances in neural information processing systems*, pages 703–711, 2014.

[10] Ryuichi Kiryo, Gang Niu, Marthinus C Du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *Advances in neural information processing systems*, pages 1675–1685, 2017.

[11] Dell Zhang and Wee Sun Lee. Learning classifiers without negative examples: A reduction approach. In *In 3rd International Conference on Digital Information Management,*, pages 638–643, 2008.

[12] Lian Yan, Robert H Dodier, Michael Mozer, and Richard H Wolniewicz. Optimizing classifier performance via an approximation to the wilcoxon-mann-whitney statistic. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 848–855, 2003.

[13] Toon Calders and Szymon Jaroszewicz. Efficient auc optimization for classification. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 42–53. Springer, 2007.

[14] Elad Eban, Mariano Schain, Alan Mackey, Ariel Gordon, Ryan Rifkin, and Gal Elidan. Scalable learning of non-decomposable objectives. In *Artificial Intelligence and Statistics*, pages 832–840. PMLR, 2017.

[15] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[16] Hyunjun Ju, Dongha Lee, Junyoung Hwang, Junghyun Namkung, and Hwanjo Yu. Pumad: Pu metric learning for anomaly detection. *Information Sciences*, 2020.

[17] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.

[18] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[19] Sumeet Shinde, Tanay Chougule, Jitender Saini, and Madhura Ingalhalikar. Hr-cam: Precise localization of pathology using multi-level learning in cnns. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 298–306. Springer, 2019.

[20] Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, and Lin Yang. Mdnet: A semantically and visually interpretable medical image diagnosis network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6428–6436, 2017.

[21] Sally Goldman and Yan Zhou. Enhancing supervised learning with unlabeled data. In *ICML*, pages 327–334. Citeseer, 2000.

[22] Xu Ji, João F Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9865–9874, 2019.

[23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[24] Daniel J Kaser and Catherine Racowsky. Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review. *Human reproduction update*, 20(5):617–631, 2014.