

乳房 X 線画像正常例で訓練した拡散モデルによる異常 検知に関する研究

Study on Anomaly Detection using Diffusion Models

Trained on Normal Mammograms

吉住 堯成*, ○三上 真理子*, 類家朝陽*, Zeng Yuwen*, 張 曉勇**, 市地
慶*, 本間 経康*

Issei Yoshizumi*, ○Mariko Mikami*, Asahi Ruike*, Yuwen Zeng*, Xiaoyang
Zhang*, Kei Ichiji*, Noriyasu Homma*

*東北大学, **仙台高等専門学校

*Tohoku University, **National Institute of Technology, Sendai College

キーワード: 異常検知 (Anomaly Detection)、乳がん (Breast cancer)、拡散モデル (Diffusion model)

連絡先: 〒980-8575 仙台市青葉区星陵 2-1 東北大学大学院医学系研究科 医用画像工学分野

吉住 堯成 Tel.: 022-717-8190, E-mail: issei.yoshizumi.p6@dc.tohoku.ac.jp

1. はじめに

乳がん検診の主流は乳房 X 線画像検査 (マンモグラフィ) を用いた画像診断であり, マンモグラフィ検診受診者数は増加している¹⁾. これに伴い, マンモグラフィ画像を読影する医師の負担増加が問題視されており, コンピュータ支援診断による読影医の補助が期待されている. 近年のコンピュータ支援診断システムは深層学習モデルを用いたものが主流である^{2),3),4)}が, このシステムで高い性能を得るためには大量の訓練データが必要となる. マンモグラフィ検診画像では病変例のデータ数が少ないため, 様々な病変の特徴を網羅的にモデルに学習させることは容易ではない. 一方で数多く存在する正常例の画像を再構成するよう訓

練したモデルにより非正常例すなわち病変を含む異常例を検知する方法がある⁵⁾. 医用画像上の異常検知にはオートエンコーダ (Autoencoder, AE)⁶⁾や敵対的生成ネットワーク (Generative Adversarial Networks, GAN)⁷⁾を用いた方法が提案されているが, それぞれ再構成画像の解像度の低さや訓練の不安定さなどの問題がある. このため従来よりも高解像度画像を再構成でき, 安定的に学習可能な深層学習モデルであるノイズ除去拡散確率モデル (Denoising Diffusion Probabilistic Model, DDPM) を用いた異常検知法も研究されている⁸⁾. DDPM はノイズを段階的に付与した入力画像からノイズを除去する過程を通じて入力画像を推定する拡散モデルであり, 脳 MRI 画像上の脳卒中や小児前腕部 X 線画像上の骨折など

異なるモダリティ上の病変を同一モデル・設定で検知可能なことが示されている⁹⁾。先行研究において高解像度 X 線画像上の異常を検知できていることから、マンモグラフィ画像上の異常検知への応用が期待される一方、乳腺など低コントラストの正常組織・構造の再構成には先行研究同様のノイズ付与設定は適さないと考えられる。本研究では、より効果的なマンモグラフィ画像診断支援に向け、DDPM に基づく手法のマンモグラフィ画像における異常検知の実現可能性と適切なノイズ設定を明らかにすることを目的とする。

2. 拡散モデルによる異常検知

2.1. 拡散モデルによる画像再構成

拡散過程では、入力画像 \mathbf{x}_0 に徐々にノイズを付し、 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ のように画像 \mathbf{x}_t を変換する。この過程は以下の近似事後分布として記述できる。

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1 - \alpha_t)\mathbf{I}) \quad (1)$$

ここで、 $\alpha_t = \prod_{s=1}^t(1 - \beta_s)$ であり、 β_t は各ステップでのノイズレベルを決める分散スケジュールを表す。ステップ t における画像 \mathbf{x}_t は、Reparameterization Trick によって \mathbf{x}_0 より次式のように直接求めることができる。

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\boldsymbol{\epsilon}, \quad (2)$$

$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

逆拡散過程では、 T ステップ分のノイズを付与した画像 \mathbf{x}_T から元の入力画像 \mathbf{x}_0 を再構成する。これは拡散過程の逆を近似するパラメータ化モデル $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ を学習することで実現される。

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_\theta(\mathbf{x}_t, t)^2\mathbf{I}) \quad (3)$$

平均 $\mu_\theta(\mathbf{x}_t, t)$ と分散 $\sigma_\theta(\mathbf{x}_t, t)^2$ は変分

下限を最小化することで学習される。逆拡散過程では、ガウス事前分布 $p(\mathbf{x}_T)$ のサンプルから始めて、 $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ を繰り返し適用することで、元のデータ分布 $p(\mathbf{x})$ に似た画像が生成される。

2.2. DDPM のノイズパラメータ

本研究では、DDPM による異常検知において、入力画像に付与されるノイズ量がモデル出力に与える影響を検討する。DDPM には付与するノイズに関する設定可能なパラメータが三つ存在する。一つ目はノイズの種類である。先行研究では平均 0, 分散 1 として生成したガウシアンノイズや複数の設定でのシンプレクスノイズが使用されている^{8),9)}。二つ目は拡散過程と逆拡散過程において、ノイズを足すステップ数の最大値を示す Inference step: T である^{8),9)}。三つ目は逆拡散過程における開始ステップの最小値を決める Noise level reconstruction (NLR) である¹⁰⁾。開始ステップとは、逆拡散過程を始めるステップ数であり、NLR から Inference step: T 間のステップ数からランダムに選ばれる。つまり、NLR と T のいずれも逆拡散過程の始まりに選ばれる画像のノイズ量を制御するパラメータである。本研究では、簡単のため、ノイズの種類はガウシアンノイズに絞ることとする。また、DDPM に基づいた医用画像向け異常検知の先行研究では共通して Inference step: $T = 1000$ と設定されていることになり、本研究においても同じ値、 $T = 1000$ を採用する。

入力画像に付与されるノイズ量に影響する三つ目のパラメータ NLR による再構成画像の違いについて考える。例えば NLR=0 の場合、逆拡散過程を始めるステップ数 t として NLR= 0 から Inference step: $T =$

1000 までのランダムな値が選択される。このとき、ステップ数は最小 $t=0$ であり、拡散過程で付与されるノイズは入力画像の画素値・信号成分に比べて極めて小さいスケールのノイズ成分となる。このため、再構成を行っても入力画像とほぼ変わらない画像が得られることが想定される。他方、 $NLR=1000$ とした場合は、逆拡散過程を開始するステップ数は最大 $t=1000$ となり、この場合には拡散過程で付与されるノイズが多いため逆拡散過程でノイズを除去しきれず、入力と全く異なる画像が再構成される可能性がある。このように、 NLR は再構成画像の特徴を大きく左右するノイズパラメータであり、その設定の違いが異常検知の性能に影響を与える可能性がある。

2.3. DDPM による異常検知

DDPM をマンモグラフィ画像上の病変の検知に応用するため、はじめに医師によって正常と診断されたマンモグラフィ画像を訓練データとして、入力画像を再構成するタスクによりモデルを訓練する。訓練を始める時点で、ノイズパラメータ NLR を設定する。訓練データには正常例のみが含まれるため、モデルは正常組織が持つ画像特徴のみを用いた画像再構成を学習することとなり、異常所見の画像特徴は学習しない。

テストデータを用いた画像再構成では、訓練済みモデルへ正常例と異常例の双方を入力する。モデルは学習した正常画像に含まれる特徴をもとに再構成を行うため、正常画像を入力した場合、入力画像に似た画像が再構成されると考えられる。一方、異常画像を入力した場合には、モデルは異常所見の画像特徴を学習していないため、正常例にみられる特徴のみからなる画像が再構成されると考えられる。なお、DDPM が出

力する再構成画像の画素値範囲は $0\sim 1$ 付近ではあるが、出力層の活性化関数の特性上、わずかに 0 より小さい画素値を持つピクセルや 1 より大きい画素値を持つピクセルが生じる場合がある。これ以降の指標値計算、および残差マップ作成に用いられる再構成画像は、画素値が $0\sim 1$ の範囲を取るように、 0 未満の画素値を 0 に、 1 より大きい画素値を 1 に置き換えたものを用いることとする。

次いで、入力画像と再構成画像の違いを視覚的に確認し、分類指標を計算するために残差マップを作成する。DDPM による再構成画像には元画像より画素値の大きいピクセルが存在する場合があり、入力画像と再構成画像の画素値の差が一部で負になる可能性があるため、先行研究に倣い、入力画像と再構成画像の絶対誤差をマップ化する。最後に、残差マップから入力画像と再構成画像の違いを算出する分類指標を算出し、算出された指標値に閾値をかけて正常と異常の分類を行う。このとき、以下に示される Youden index J が最大となる閾値を採用する。

$$J = \text{Sensitivity} + \text{Specificity} - 1 \quad (4)$$

ここで、Sensitivity は感度、Specificity は特異度である。

3. 検証実験

3.1. データセット

本研究では、実験用のデータセットとして、宮城県対がん協会 (Miyagi Cancer Society, MCS) より提供された検診マンモグラフィデータのうち、デジタルマンモグラフィ装置で撮影され、両側の画像が含まれている症例のうち撮影方向が頭尾 (Cranio-Caudal, CC) 軸方向の画像に限定して使用

した。このデータには、全症例に対して医師の読影に基づいた BI-RADS (Breast Imaging Reporting and Data System) 分類が付してある。また、異常例に対しては、生検の結果をもとに良悪性のラベルが付けられており、東北大学病院で勤務経験のある医師の協力のもと、画像上の病変位置を示すアノテーションが付与されている。

MCS データに含まれるデジタルマンモグラフィ画像は画像サイズ (2294, 1914) であるが、計算機メモリの都合により、画像を縮小してモデルに入力する必要があったため、2通りの前処理を適用した。一つ目の前処理は画像全体をそのまま縮小する方式であり、画像サイズ (2294, 1914) の原画像を画像サイズ (160, 160) へとリサンプリングした。この方式では、もともとのマンモグラフィ画像で小さく描出されている構造は潰れてしまい、モデルが小さい構造物の特徴を学習できない可能性がある。モデルの訓練に使用する全体画像は、訓練データとして割り当てた 676 症例における左右乳房の CC 方向の画像であり、計 1352 枚である。バリデーションデータについても同様に、84 症例から計 168 枚を使用した。二つ目の前処理の方式は、画像のパッチ分割である。画像サイズ (2294, 1914) のデジタルマンモグラフィ画像から、MCS データに含まれる最も大きい腫瘍がパッチ内に収まるようなサイズ (320, 320) のパッチを取り出す。このとき、乳房領域外だけが映った黒画素が大部分を占めるパッチが出ないように、4割以上の画素が乳房領域を含むようなパッチを対象とする。取り出したパッチサイズ (320, 320) のパッチをパッチサイズ (160, 160) にリサンプリングにより縮小した。パッチ分割方式では訓練に要する計算時間の

都合上、モデルの訓練に使用するパッチを 5000 枚に限定した。訓練データとして割り当てた 676 症例 69197 枚から、必ずすべての症例から 1 枚以上、1 症例あたり平均 7, 8 枚程度選ばれるよう、ランダムに 5000 パッチを取り出した。バリデーションデータについても同様に、84 症例 8629 枚から合計 500 パッチをランダムに取り出した。

3.2. 分類指標

分類指標には、病変部位で残差マップの値が大きくなると考えられることから、平均絶対誤差 (mean absolute error, MAE) を用いた。なお、 $I(i, j)$ は入力画像の画素値、 $\hat{I}(i, j)$ は再構成画像の画素値、 (i, j) は画像中の画素位置、 M は画像の高さ、 N は画像の幅を意味する。

$$MAE = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N |I(i, j) - \hat{I}(i, j)| \quad (5)$$

3.3. 評価指標

評価指標には、受信者動作特性曲線 (Receiver Operating Characteristic curve, ROC 曲線) と Area Under the Curve (AUC) を用いた。ROC 曲線は縦軸に真陽性率、横軸に偽陽性率をとった曲線である。AUC は ROC 曲線の下部の面積であり、値が 1 に近いほど分類性能が高いとされる。

3.4. 比較対象モデル

比較対象モデルとして、マンモグラフィのパッチ画像で石灰化向けに異常検知を行った先行研究⁵⁾より提案されているオートエンコーダを用いる。オートエンコーダはエンコーダとデコーダから構成されるニューラルネットワークアーキテクチャであり、エンコーダで入力データの特徴を圧縮し、圧縮された特徴をデコーダで復元することで入力と同じような入力データを再構築するモデルである。本研究では先行研究をも

とにエンコーダは画像サイズ (160, 160) を入力として受け付けるモデルを実装・構築した。本モデルではエンコーダにおいて畳み込み層と最大プーリング層を組み合わせ、3段階の特徴圧縮を行い最終的に潜在空間までに入力画像を (20, 20, 1) の特徴量へと圧縮する。デコーダでは、潜在空間に圧縮された (20, 20, 1) の特徴量から、畳み込み層とアップサンプリング層を通して入力画像の再構成を行う。

3.5. 実験設定

DDPM のノイズパラメータによる異常検出性能の違いを確かめるため、NLR を 0 から 400 まで 50 ずつ変化させたモデルを訓練・評価した。また、すべてのノイズパラメータにおいて、マンモグラフィ画像全体とパッチ分割した画像の2つのデータセットを試行した。簡単のため、DDPM の訓練におけるシード値を固定し、訓練中の訓練データの選択パターンとノイズの載せ方を一定とした。なお、評価においては、DDPM でのノイズに関するシード値は固定せず、設定したパラメータの範囲内でランダムにノイズ付与されるものとした。ノイズの載せ方の違いによる画像再構成のばらつきを考慮に入れるため、画像再構成を 10 回試行した。比較対象であるオートエンコーダも同様の2つのデータセットで訓練・評価した。

3.6. 結果

全体画像およびパッチ画像を対象とした再構成画像と残差マップの例を図1, 図2にそれぞれ示す。図1の全体画像の場合、NLR = 0 で、入力画像に似た画像が再構成され、残差は比較的小さい。また、残差マップはガウシアンノイズを乳房領域の形に切り取ったような画像になった。NLR が上がると、

入力画像の乳房の形状とは異なる構造物が再構成される領域が広がり、それに伴って残差マップも入力画像にはない構造を強調するような画像となった。図2のパッチ画像の場合、NLR = 0 で入力画像と再構成画像は類似しており、微細な乳腺構造まで再構成できている。残差マップは全体的にガウシアンノイズを載せたような画像である。NLR が上がるにつれて、入力画像にはない乳腺のような構造が増え、再構成された乳腺構造の高輝度領域に沿うような残差分布がみられる。

全体画像を対象とした各モデルによる ROC 曲線とその AUC を図3, 図4にそれぞれ示す。図3の結果から、ノイズパラメータが比較的低い DDPM とオートエンコーダによる分類性能が高かったことが読み取れる。また、NLR が 200 以上の場合、NLR が高いほど分類性能が低くなった。図4の結果から、NLR = 400 以外は AUC : 0.80 を超え、特に NLR = 0, 200 が今回作成したモデルの中で最高の AUC : 0.86 を示した。

3.7. 考察

画像種類による再構成画像、残差マップ、異常検出性能の違いについて述べる。全体画像では、入力前に病変特徴が潰れていたため、病変位置特定・正常と異常の分類は困難であった一方で、パッチ画像では、正常乳腺構造・病変の画像特徴を残していたため、異常検知に有効であった。このことから、ある程度、もとの乳房 X 線画像の特徴を残した解像度が必要だと考えられる。

NLR による再構成画像、残差マップ、異常検知性能の違いについて述べる。パッチ画像を用いた低 NLR の DDPM で分

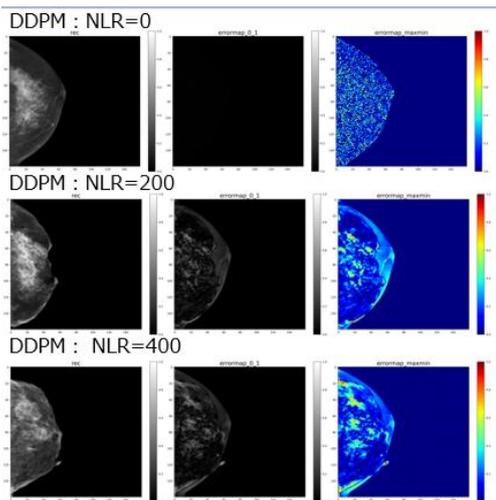


図 1. 全体画像における再構成画像と残差マップ

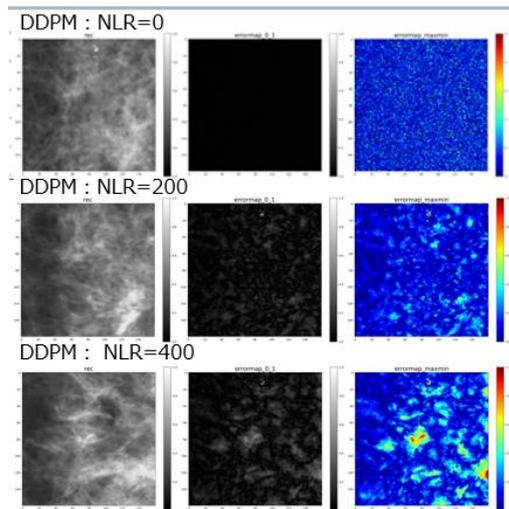


図 2. パッチ画像における再構成画像と残差マップ

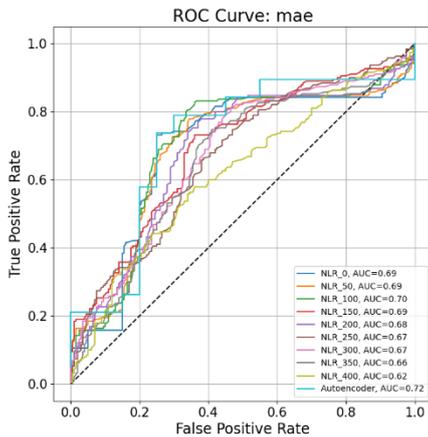


図 3. 全体画像を対象とした ROC 曲線と AUC

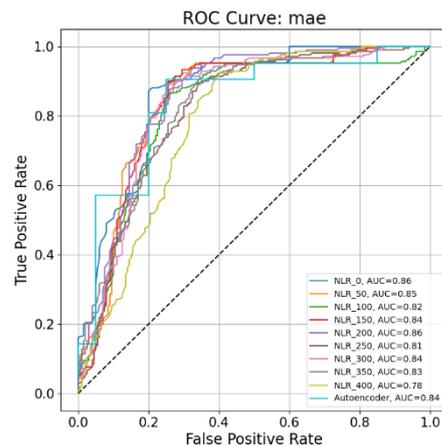


図 4. パッチ画像を対象とした ROC 曲線と AUC

類性能が高いことが分かった。これは、画像再構成において、病変部の特徴をおおまかに消し、正常部の特徴は元画像に近い状態で再構成されたため、相対的に病変部分の残差が際立ち、区別できたと考えられる。

本研究では、少ないデータセットでの評価となり、訓練とテストでデータの分布の偏りが結果に影響した可能性がある。したがって、画像数・症例数を増やした検証が望まれる。また、本研究では NLR の

み検討したが、マンモグラフィ画像上の異常検知における他のパラメータの影響は未検討である。そのため、他のパラメータが異常検知性能に与える影響についても比較・検討が必要である。さらに、本研究では腫瘍と石灰化をまとめて扱ったが、病変ごとに画像特徴が異なるため、適した分類指標も異なる可能性がある。したがって、病変の種類ごとに適切な分類手法を検討することが重要である。

4. おわりに

本研究では、乳がん検診向けの読影支援のために、乳房 X 線画像（マンモグラフィ画像）上の病変検出への拡散モデルを用いた異常検知の応用可能性について検討を行った。実験の結果より、他の医用画像上の異常検知で有効であった設定よりも低いノイズパラメータを用い、パッチ画像を使用した場合に、拡散モデルは、従来の異常検知向けモデルであるオートエンコーダを超える異常検知性能が確認された。一方、実用化に向けては、パッチ画像における評価方法の見直しや、他の適切なノイズ種類・パラメータの探索などの検討がさらに必要である。

参考文献

- 1) ‘国立がん研究センターがん情報サービス「がん登録・統計」 | 国民生活 基礎調査（厚生労働省大臣官房統計情報部）’. Accessed: Jan. 13, 2025. [Online]. Available: https://ganjoho.jp/reg_stat/statistics/stat/screening/dl_screening.html#a18
- 2) C. R. Taylor, N. Monga, C. Johnson, J. R. Hawley, and M. Patel, ‘Artificial Intelligence Applications in Breast Imaging: Current Status and Future Directions’, *Diagnostics* 2023, Vol. 13, Page 2041, vol. 13, no. 12, p. 2041, Jun. 2023, doi: 10.3390/DIAGNOSTICS13122041.
- 3) H. Fujita, ‘AI-based computer-aided diagnosis (AI-CAD): the latest review to read first’, *Radiol Phys Technol*, vol. 13, no. 1, pp. 6–19, Mar. 2020, doi: 10.1007/S12194-019-00552-4/FIGURES/11.
- 4) H. P. Chan, R. K. Samala, and L. M. Hadjiiski, ‘CAD and AI for breast cancer-recent development and challenges’, *Br J Radiol*, vol. 93, no. 1108, Apr. 2020, doi: 10.1259/BJR.20190580.
- 5) R. Hou et al., ‘Anomaly Detection of Calcifications in Mammography Based on 11,000 Negative Cases’, *IEEE Trans Biomed Eng*, vol. 69, no. 5, pp. 1639–1650, May 2022, doi: 10.1109/TBME.2021.3126281.
- 6) D. Bank, N. Koenigstein, and R. Giryes, ‘Autoencoders’, *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Third Edition*, pp. 353–374, Mar. 2020, doi: 10.1007/978-3-031-24628-9_16.
- 7) I. J. Goodfellow et al., ‘Generative Adversarial Networks’, *Sci Robot*, vol. 3, no. January, pp. 2672–2680, Jun. 2014, Accessed: Dec. 06, 2024. [Online]. Available: <https://arxiv.org/abs/1406.2661v1>
- 8) J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, ‘DDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise’, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2022-June, pp. 649–655, 2022, doi: 10.1109/CVPRW56347.2022.00080.
- 9) C. I. Bercea, B. Wiestler, D. Rueckert, and J. A. Schnabel, ‘Diffusion Models 58 with Implicit Guidance for Medical Anomaly Detection’, Mar. 2024, Accessed: Nov. 28, <http://arxiv.org/abs/2403.08464>

10) Cosmin Bercea • Benedikt Wiestler • Daniel Rueckert • Julia A. Schnabel, 'GitHub - compai-lab/2024-miccai-bercea-thor'. Accessed: Jan. 27, 2025. [Online]. Available: <https://github.com/compai-lab/2024-miccai-bercea-thor>