

ロットごとのウェハマップの重ね合わせによる欠陥パターンの顕在化とその分類

中村 陸人[†] 永村 美一[†] 新井 雅之^{††} 福本 聡[†]

[†] 東京都立大学大学院 システムデザイン研究科 情報科学域 〒191-0065 東京都日野市旭が丘 6-6

^{††} 日本大学 生産工学部 〒275-8575 千葉県習志野市泉町 1-2-1

E-mail: [†]nakamura-rikuto@ed.tmu.ac.jp, [†]yoshikazu.nagamura.vj@renesas.com, ^{††}arai.masayuki@nihon-u.ac.jp,
[†]s-fuku@tmu.ac.jp

あらまし 従来のウェハマップ欠陥分類手法は単一ウェハマップを対象としている。しかし、分類で重要となるシステムチック欠陥は製造ロット単位で発生する可能性がある。著者らの知る限り、これまでにそのような欠陥を高精度かつ効率的に検出・分類する手法は見当たらない。本研究では、ロット単位のウェハマップ重ね合わせによる潜在的な欠陥情報の抽出手法を提案する。重ね合わせによって、ノイズに埋もれていた異常を可視化し、新たな欠陥パターンを発見した。さらに、重ね合わせマップ特有の欠陥パターンを高い精度で分類・検知することを目標とし、本稿ではその第一歩として、空間認識に優れる CapsNet を用いた実験結果について報告する。

キーワード ウェハマップ, ロット, 重ね合わせ, CapsNet, 画像分類, 異常検知, 半導体, AI

Rikuto NAKAMURA[†], Yoshikazu NAGAMURA[†], Masayuki ARAI^{††}, and Satoshi FUKUMOTO[†]

[†] Department of Computer Science, Graduate School of Systems Design, Tokyo Metropolitan University 6-6 Asahigaoka, Hino-shi, Tokyo, 191-0065 Japan

^{††} College of Industrial Technology, Nihon University 1-2-1 Izumi-cho, Narashino-shi, Chiba, 275-8575, Japan

E-mail: [†]nakamura-rikuto@ed.tmu.ac.jp, [†]yoshikazu.nagamura.vj@renesas.com, ^{††}arai.masayuki@nihon-u.ac.jp,
[†]s-fuku@tmu.ac.jp

1. はじめに

半導体は各種インフラの構築に不可欠なものであり、日常生活における多様なシステムにも広く使用されている。そのため半導体チップの需要は拡大しており、これに伴い製造における品質基準や生産効率に対する要求は一層厳格化している。したがって、製造工程における歩留りの向上は、現代の半導体産業における最重要課題の一つである。

歩留りを向上させるためには、不良チップを製造過程の早期段階において効率的かつ正確に検出することが求められる。製造現場で日々得られる検査データの情報は大量かつ多様であり、不良チップの解析に向けて様々な手法が開発されてきた。近年では、AIを用いた分類によってウェハマップ上の欠陥パターンを解析する手法が注目されている。

ウェハマップとは、半導体ウェハ上のダイ（切り出し前のLSI）に対する様々な測定結果の二次元分布を画像化したものである。各ダイに対する製造テストの結果（良品/不良品）をプロットしたウェハマップ上の模様（欠陥パターン）を解析す

ることで、半導体の不良原因となる製造装置や工程の特定に有効な情報を得ることができる。欠陥には、一定の確率で偶発的に発生するランダム欠陥と、各種の装置異常などに起因するシステムティック欠陥がある。ウェハマップの解析において重要となるのは後者である。ただし、従来のウェハマップ解析では、得られる欠陥情報が個別のウェハ単位に限定され、製造ロット単位で内在する欠陥情報が必ずしも顕在化しないという課題がある。

実際の半導体製造現場における歩留り管理や品質検査においては、複数の検査結果を統合して欠陥の傾向を確認する手法が用いられている。例えば、Madge ら [1] は、コンポジットマップ（“Composite Map”）と呼ばれる手法を検証している。コンポジットマップとは、電氣的テストにおいて不良と判定されたダイの位置情報を、複数のウェハにわたって統合しマッピングしたものである。具体的には、ウェハ上の各座標において発生した不良ダイの総数を蓄積し、その数に比例した大きさの正方形を該当位置に描画する。これによって、どの位置で系統的に不良が発生しているかを視覚化している。

しかしながら、これはあくまで技術者による視覚的な確認を補助する手法にとどまっている。機械学習や深層学習を用いた欠陥パターンの自動分類研究においては、依然として単一のウェハマップを対象とした解析が主流であり、ウェハマップの重ね合わせ情報を AI モデルの学習に直接組み込むアプローチは見当たらない。

本研究では、ウェハマップ解析の新たな手法を提案する。基本的なアイデアは、複数のウェハマップを重ね合わせることで新たなマップを生成し、隠れた欠陥情報を抽出することである。具体的には、ひとつの製造ロットに含まれる 25 枚のウェハマップを統合してひとつのマップを生成し、ロットごとの欠陥情報を表す新たな正解ラベルを付与する。生成したデータセットと画像認識モデルのひとつである CapsuleNetwork(CapsNet)を用いて、正常、異常の判別、欠陥種別の特定を目的とした教師あり分類、ならびに正常データだけを用いた異常検知の実験を行い、提案手法の有効性を検証する。

本稿の構成は以下の通りである。2. では、ウェハマップおよび CapsNet の構造に関する事前知識について説明する。3. では、本研究における提案手法の概要について詳述する。4. では、構築したデータセットに対する実験結果と考察を述べる。5. で本稿をまとめ、今後の課題を述べる。

2. 背景

本節では、本研究に関連するウェハマップ、CapsNet の概要について述べる。

2.1 ウェハマップ

半導体製造工程では、シリコンウェハ上に回路を形成し、ダイと呼ばれる半導体チップを格子状に配置する。製造プロセスの後、個々のダイに対してプロービングテストと呼ばれる電氣的検査が実施され、正常か欠陥かを判定する。この検査結果に基づいて、ウェハ上のダイの空間的な配置と正常・欠陥の状態を可視化した画像データがウェハマップである。

本研究では、実際の半導体製造現場で収集されたウェハマップデータセットである WM-811K [2], [3] を使用する。本データセットは 811,457 枚のウェハマップから成り、そのうち 172,950 枚には、欠陥の形状パターンに基づいてラベルが付与されている。具体的には、None, Center, Donut, Edge-Loc, Edge-Ring, Loc, Near-full, Scratch, Random の 9 種類である。このうち None は、ランダム欠陥だけを含む状態を表し、それ以外の 8 種類がシステムティック欠陥を含む状態を表す。図 1 に各ラベルに対応するウェハマップの例を示す。

WM-811K の各ウェハマップにはロットと呼ばれる製造単位の情報が含まれている。本データセットは、46,393 ロットのうち約 47% のロットが、1 ロットあたりの最大枚数である 25 枚のウェハマップで構成されている。

2.2 CapsNet

CapsNet (Capsule Network) は、Sabour ら [4] が提案したニューラルネットワークモデルである。従来の畳み込みニューラルネットワーク (CNN) は、プーリング層を用いて特徴マップをダウンサンプリングすることで、画像特徴を集約し、分類の汎

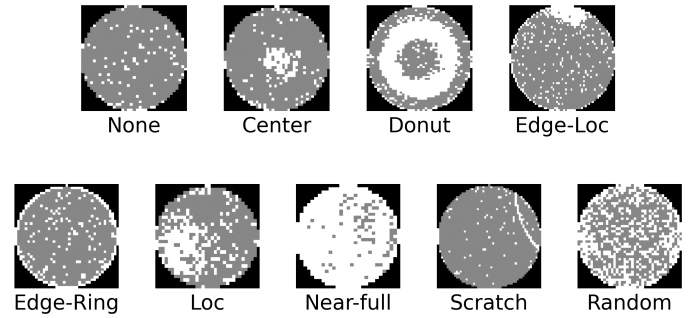


図 1 WM-811K の欠陥パターン例

用性を向上させる。しかし、プーリング層による情報の欠落は、特徴間の空間的な位置情報を失わせるという課題がある。

これに対して CapsNet は、スカラー値ではなくベクトルを出力として特徴を扱う。プーリング層の代わりにダイナミックルーティング (dynamic routing) と呼ばれる手法を用いて、空間的な特徴を抽出している。CapsNet は、特徴抽出及び分類を行う分類器と、と入力画像を復元するデコーダの二つのネットワークから成り、ダイナミックルーティングは分類器で使用されている。

3. 提案手法

本節では、重ね合わせ画像における正解クラス作成手法と、CapsNet を用いた分類および異常検知の手法について述べる。

3.1 正解クラスの作成

A. 対象ロットの選定と重ね合わせ画像の生成

WM-811K に含まれるロットはウェハ枚数が不均一であるため、前処理としてロットを選定した。本研究では、1 ロットあたりの最大構成枚数である 25 枚のウェハマップを有するロットだけを選定し、重ね合わせの対象とした。少数枚数のウェハマップで構成されるロットは、試作ロットや搬送ミスによる抜き取り、あるいは記録の中断が行われたロットである可能性が高く、装置異常の特定には不向きであるためである。

選定したロット内の各ウェハマップに対して、全ウェハマップの平均サイズを考慮したリサイズを行った。リサイズ後、同一ロット内の 25 枚のウェハマップから 1 枚の重ね合わせ画像を生成した。

B. クラスタリング手法

生成した重ね合わせ画像には、ラベルが存在しない。そこで本研究では、パターンを分類し独自の正解ラベルを作成するために、教師なし学習であるクラスタリングを用いた。クラスタリングとは、データの特徴量から類似性を計算し、似たデータをグループ分けする手法である。なお、本研究ではクラスタリングの結果生成されたグループを「クラスタ」と呼び、それらの特徴に基づいて人間がラベル付けを行ったものを「クラス」と呼称して区別する。

図 2 に、本研究における正解クラス作成のフローチャートを示す。クラスタリングは 2 段階で実施した。

第 1 段階では、ロット内の各ウェハに付与されているラベル情報の構成比率に基づいてクラスタリングし、全 21,762 ロッ

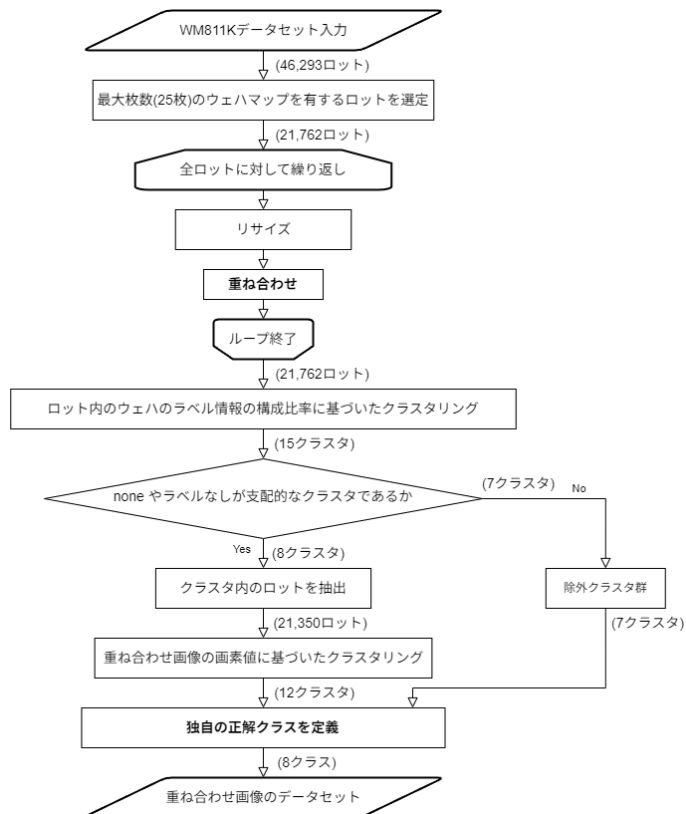


図2 正解クラス作成手順

トを 15 のクラスに分割した。これによって、既存のシステムティック欠陥ラベルが支配的なロットをだまかに分類し、第 2 段階目のクラスターリング対象から除外した。

第 2 段階では、第 1 段階の分類結果のうち、None およびラベルなしのウェハマップが支配的なクラス（8 クラス、21,350 ロット）に着目した。これらのクラス内に含まれるロットを抽出し、重ね合わせ画像の画素値に基づいたクラスターリングを実施した。その結果、全 21,350 ロットを 12 のクラスに分割した。

この 2 段階の手順を踏むことで、既存ラベルで判別可能な顕著なロット異常を分離しながら、ラベル付けが漏れていた、もしくは None のウェハが多く含まれるロットから潜在的な異常パターンを抽出することが可能となる。最終的に、画素値ベースのクラスターリング結果と、第 1 段階で除外したクラス群を統合し、WM-811K の元の名称とは異なる独自の 8 クラスを定義した。本研究では、この独自のクラス群を以降の実験におけるデータセットとして使用した。

3.2 実験設定

A. 正常、異常クラスの定義

前節のクラスターリング結果に基づいて、重ね合わせマップにおけるシステムチックな欠陥パターンが顕著なクラスを「異常」と定義し、重ね合わせマップにおけるランダム欠陥だけを含んだクラスを「正常」と定義する。本研究では、この定義に基づいて実験する。

B. 教師あり分類

CapsNet が重ね合わせ画像の欠陥パターンを識別できるかど

うかを検証するため、以下の二つの教師あり分類実験を実施した。また、少数枚数のデータしか持たないクラスの学習不足を防ぐため、学習用データに対してだけ、画像の回転および反転処理のデータ拡張を実施し、各クラスの学習データ数を調整した。

- **正常、異常の 2 クラス分類：**全データセットを用いて、入力データが正常か異常かを判別する。これによって、モデルの基礎的な分離性能を検証する。

- **異常データの分類：**異常と定義されたデータだけを用い、詳細な欠陥種別を特定する。これによって、異常データ同士の微細な特徴差を識別する多クラス識別能力を検証する。

C. 正常データ学習による異常検知

未知の欠陥に対する検知能力を検証するため、教師なし学習のアプローチをとる。具体的には、正常データだけで CapsNet を学習させ、入力画像を再構成させる。テスト時には、未学習である異常データも入力し、入力画像と再構成画像との画素値の平均二乗誤差を算出する。正常データだけで最適化されたモデルは、未知の異常パターンを正しく再構成できないため、誤差が大きくなる。この誤差を異常度の指標とし、閾値を設けることで正常と異常を区別した。

D. 評価指標

分類結果、および上述の閾値によって判定された異常検知結果と実際の正解ラベルとの整合性を評価するため、正解率 (Accuracy)、再現率 (Recall)、適合率 (Precision)、F1 スコア (F1-score) の四つの指標を用いた。正解率は全データのうち正しく分類できた割合を示す。再現率は特定のラベルを持つデータ全体に対して正しく分類できた割合であり、欠陥の見逃しの少なさを示す。適合率は、特定のラベルに分類されたデータ全体に対して、実際にそのラベルであった割合であり、過検出の少なさを示す。F1 スコアは、再現率と適合率の調和平均であり、それらのバランスを取った指標を示す。

これらの指標のうち、正解率を除く 3 種類については、すべての欠陥パターンに対して、個別で算出することで、欠陥パターンの種類ごとの精度を比較する。また、学習したモデルの全体を評価する指標として、欠陥パターンごとの評価指標の平均を求める必要がある。本研究では、テスト用画像枚数の差を考慮しない重みなし平均 (macro avg) と、枚数の差を考慮する重みあり平均 (weighted avg) の 2 種類を計算して比較した。

さらに異常検知においては、閾値を変動させた際のモデルの検知性能の推移を評価するため、ROC 曲線 (Receiver Operating Characteristic curve) および AUC (Area Under the Curve) を用いた。ROC 曲線は、閾値を連続的に変化させた際の真陽性率 (Sensitivity: 実際の異常データのうち、正しく異常と検知できた割合、すなわち上記の分類における再現率と同義) と偽陽性率 ($1 - \text{Specificity}$: 実際の正常データのうち、誤って異常と判定されたデータの割合) をプロットしたものであり、AUC はその曲線の下側の面積を表す。AUC は 1 に近いほど検知性能が高いことを示す総合的な指標である。

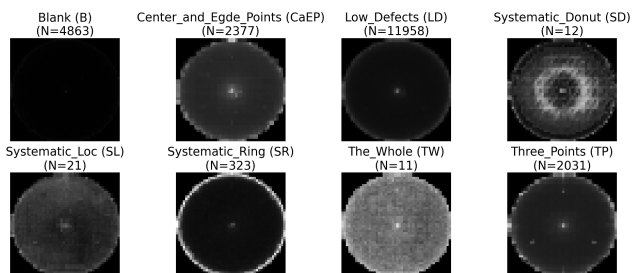


図3 統合・定義された独自の8クラスの平均画像

表1 2クラス分類結果

label	Recall	Precision	F1-score	Support
Normal	0.9979	0.9932	0.9956	955
Abnormal	0.9579	0.9925	0.9842	3365
macro ave	0.9869	0.9929	0.9899	4320
Weighted ave	0.9931	0.9931	0.9930	4320
Accuracy	0.9931			

4. 実験結果および考察

4.1 正解クラス

図3に、各クラスに属する重ね合わせマップの平均画像と、ロット数を示す。特筆すべき結果は、WM-811Kの既存の9種類のラベルには存在しない欠陥パターンである“Three.Points”クラスが顕在化したことである。このクラスを構成するウェハの96.6%はラベルがついていないウェハであった。

Three.Pointsクラスには、欠陥分布が三角形型に現れるという特徴がある。これは、製造装置内でウェハを固定・搬送するための支持棒などが接触することによって発生したものと考ええる。

さらに図4に、Three.Pointsクラスに分類されたあるロットの単一ウェハマップ25枚の一覧を示す。この図から明らかにように、ウェハを1枚ずつ確認しただけでは、ノイズに紛れて3点の欠陥パターンを視認・判定することは困難である。重ね合わせ処理は、こうした偶発的ではない装置固有の癖を可視化・特定する上で非常に有効であると考ええる。また、多くのクラスにおいてウェハの中心部と円周部に微細な欠陥が発生していることも分かる。

本研究では、前述した微細な欠陥とは異なるシステマチックな欠陥パターンが顕著な六つのクラス（Center_and_Edge_Points, Systematic_Donut, Systematic_Loc, Systematic_Ring, The_Whole, Three.Points）を「異常（Abnormal）」と定義した。一方、それ以外の二つのクラス（Blank, Low_Defects）を「正常（Normal）」と定義した。

4.2 教師あり分類

正常、異常の2クラス分類結果を表1に示す。正解率が99.31%という極めて高い精度を達成し、提案手法が正常と異常を明確に分離可能であることが示された。

異常データだけをを用いた6クラス分類結果を表2に示す。正解率は96.86%と高い精度が得られた。特にThree.Pointsは高い精度で識別できており、CapsNetが局所的な配置特徴を捉

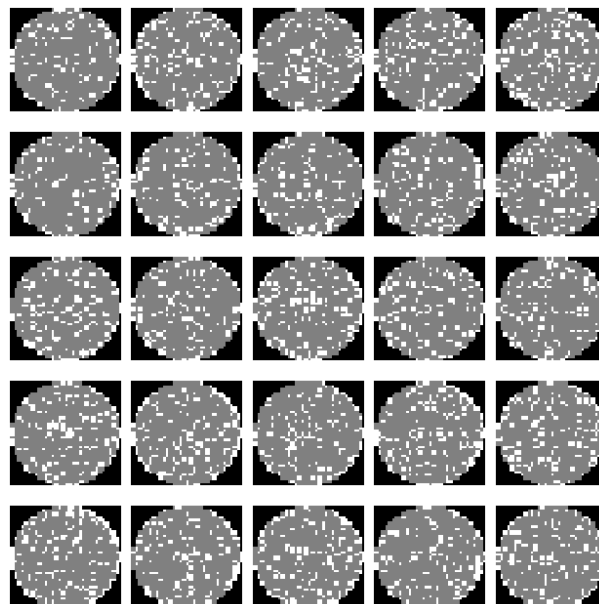


図4 Three.Pointsに分類されたロットの単一ウェハマップ一覧

表2 6クラス分類結果

label	Recall	Precision	F1-score	Support
CaEP	0.9853	0.9874	0.9863	476
SD	1.0000	1.0000	1.0000	2
SL	0.5000	0.4000	0.4444	4
SR	0.8769	0.9048	0.8906	65
TW	0.5000	0.5000	0.5000	2
TP	0.9704	0.9657	0.9681	406
macro ave	0.8054	0.7930	0.7982	955
Weighted ave	0.9686	0.9691	0.9688	955
Accuracy	0.9686			

えていることがわかる。一方で、データ拡張を行ってもなお少数クラスの誤分類が見られ、データの個数に関係なく、全てのクラスを平等に扱って計算したmacro average F1スコアは0.7982に留まった。これは、クラス間のデータ数の不均衡を示している。

4.3 正常データ学習による異常検知結果

図5は、テストデータにおける異常検知性能を示すROC曲線である。AUCは0.965となっており、異常検知モデルが十分に機能していると考えられる。

図6に、テストデータにおける入力画像と再構成画像間の誤差の分布を示すヒストグラムを示す。本グラフの横軸は誤差を、縦軸はデータの密度を表している。正常データだけを学習したモデルにとって、未知の欠陥パターンを持つ異常データは正確な再構成が困難であるため、異常データの誤差は正常データと比較して高誤差側へ分布する傾向がある。グラフから、正常データの再構成誤差は閾値以下の狭い範囲に集中して分布しているのに対して、異常データの誤差は閾値を超えて広く分布しており、両者が閾値を境界にして良好に分離できていることが確認できる。

正解ラベルと照らし合わせた検知結果を表3に示す。正解

5. む す び

本研究では、ロット単位のウェハマップ重ね合わせと CapsNet を用いた分類および異常検知手法を提案した。実験の結果、既存ラベルにない新たな欠陥パターン Three_Points を発見し、提案手法が潜在的な装置異常の可視化に有効であることを示した。さらに、正常データだけを学習させた異常検知においては、異常データにおける再現率が 95.31% に達し、致命的な欠陥の見逃しを防ぐ実用的なモデルが構築可能であることを示した。

今後の課題として、第 1 に、より良いクラスタリング手法の模索や、画像生成 AI 等を活用したデータ拡張の検討による、クラス間のデータ不均衡の解消が挙げられる。第 2 に、CapsNet におけるパラメータの最適化や、他のモデルを用いた CapsNet との対照実験を行うことが挙げられる。

文 献

- [1] R. Madge, M. Rehani, K. Cota and W. R. Daasch, “Statistical post-processing at wafersort-an alternative to burn-in and a manufacturable solution to test limit setting for sub-micron technologies,” Proceedings 20th IEEE VLSI Test Symposium, Aug. 2002.
- [2] Kaggle, WM811K wafermap, Accessed Jan.29, 2026, [Online]. Available: <https://www.kaggle.com/datasets/qingyi/wm811k-wafer-map>
- [3] M.-J. Wu, J.-S. R. Jang and J.-L. Chen, “Wafer map failure pattern recognition and similarity ranking for large-scale data sets,” IEEE Transactions on Semiconductor Manufacturing, vol. 28, no. 1, pp. 1-12, Feb. 2015.
- [4] S. Sabour, N. Frosst and G. E. Hinton, “Dynamic Routing between Capsules,” Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), pp. 3859-3869, Dec. 2017.

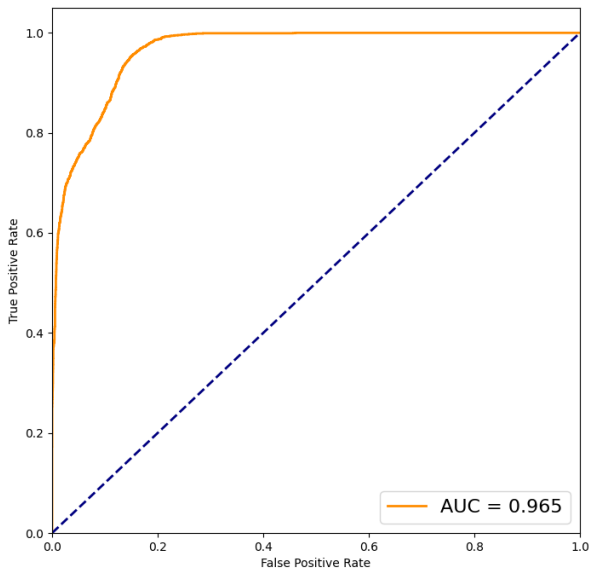


図 5 ROC 曲 線

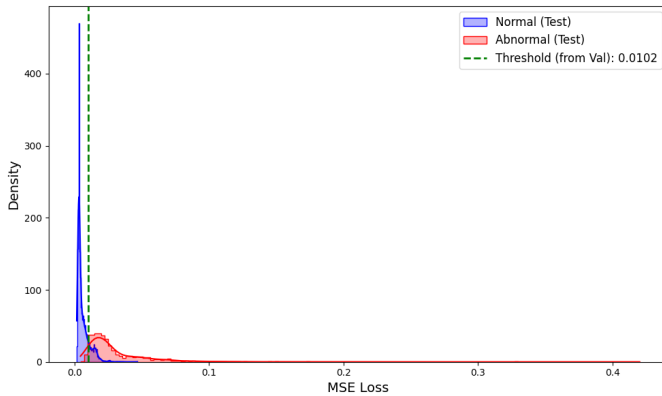


図 6 入力画像と再構成画像間の誤差の分布と閾値

表 3 異常検知結果

label	Recall	Precision	F1-score	Support
Normal	0.8505	0.9629	0.9032	5457
Abnormal	0.9531	0.8169	0.8798	3820
macro ave	0.9018	0.8899	0.8915	9277
Weighted ave	0.8927	0.9028	0.8935	9277
Accuracy	0.8927			

率は 89.27% を記録した。異常クラスの再現率は 0.9531 と高い値を示している。これは、実際の異常データの 95% 以上を正しく異常として検知できていることを意味する。一方、適合率は 0.8169 と相対的に低い値となった。これは、正常データの一部を誤って異常と判定する「過検出」が発生していることを示している。これらは、実用化の観点からは不十分な水準であると考えられ、さらなる改善が求められる。