

# 眼底画像マルチラベル分類モデルの ロバストネス特性の評価と改良

松本大蔵<sup>1</sup>, 宇都翼<sup>2</sup>, 伊野田悟<sup>1</sup>, 近藤佑亮<sup>2</sup>, 劉弘毅<sup>3</sup>, 東洋介<sup>3</sup>, 高橋宏典<sup>1</sup>, 高山卓也<sup>1</sup>  
反保宏信<sup>1</sup>, 橋本悠人<sup>1</sup>, 長岡広祐<sup>1</sup>, 高橋秀徳<sup>1,2</sup>, 川島秀俊<sup>1</sup>, 柳靖雄<sup>2,4</sup>

1. 自治医科大学眼科学講座, 2. DeepEyeVision株式会社
3. 株式会社Citadel AI, 4. 横浜市立大学視覚再生外科

# 栃木県眼科学会 利益相反開示

筆頭演者：松本 大蔵

高橋秀徳：【I】DeepEyeVision株式会社、【P】

# 緒言

- 2018年、2020年にはIDx-DR®やEyeArtなどの糖尿病網膜症AI自動診断システムがFDAによって承認された。

Abramoff MD, et al. *NPJ Digit. Med.* 2018

- 眼底画像の読影におけるAIの補助によって、眼科診断のさらなる効率化が試みられている。
- AIに入力される眼底写真は画質の状態が良いものばかりとは限らず、それが軽度な外乱(画像の明るさ、暗さ、歪みなど)でも大きく影響を受ける問題がある。

# 緒言

- 黄斑変性が見られる画像
- AIは、オリジナル画像については正しく分類できたが、ぼかしを掛けると、疾患がないと判断した。



オリジナル画像



僅かなぼかしを掛けた画像

# 緒言

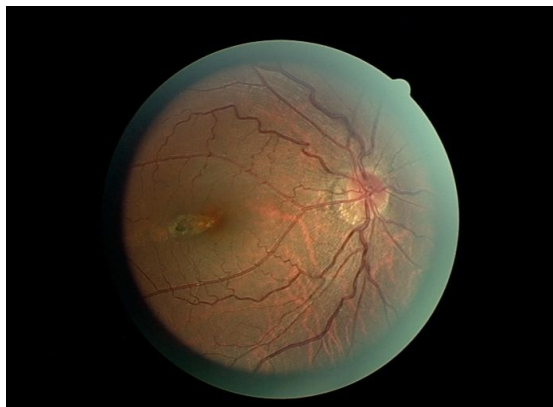
- 構築したAIがどのような外乱に対して弱いのかを調べることは世の中にAIを出す上で重要となってくる。
- 眼底画像への外乱がAIの性能に及ぶ影響について明確に示した研究はない。

# 緒言

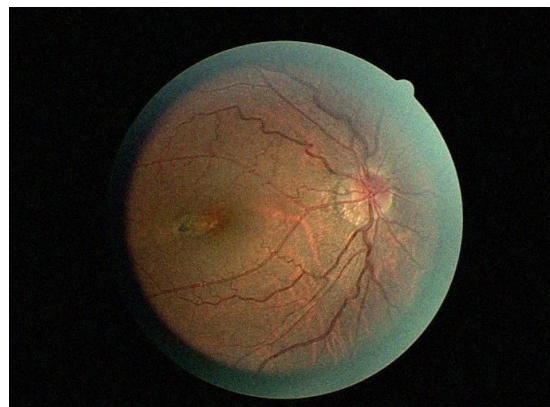
- 一方、眼底画像を読影する際に、撮影時の画像に対する外部からの干渉、外乱が問題となってくる。
- 眼底画像への外乱の例
  - カメラ内部で発生するノイズ  
(コントラストやディストーション)
  - カメラ外部の撮影環境から発生するノイズ(明るさの変動)
  - カメラの位置変動によって発生するノイズ  
(眼底の大きさ、位置の変動)

# 緒言

- S1、黄斑変性、眼底血管蛇行が見られる画像
- AIは、オリジナル画像については正しく分類できたが、ガウス雑音を足すと、これらを全て分類できなかった。



オリジナル画像



ガウス雑音を足した画像

# 目的

- 眼底画像内の疾患を推定するマルチラベル分類AIのロバストネス特性(=入力への外乱に対して、AIがどれだけ頑強か)を検証する。
- それを踏まえた改良施策を実施し、特性変化を評価する。



# 方法

- 自治医科大学で収集した通常画角カラー眼底写真を使用
  - 約40万枚、50疾患
  - このうち、疾患が見られない画像は約27万枚

# 方法

- 自治医科大学で収集した  
通常画角カラー眼底写真を使用

疾患	画像枚数
S1※	50005
H1※	34799
S2※	13442
緑内障	8380
H2※	5411
ドルーゼン	3421
白内障	3334
網膜前膜	3053
視神経乳頭陥凹	2364
糖尿病網膜症	2241

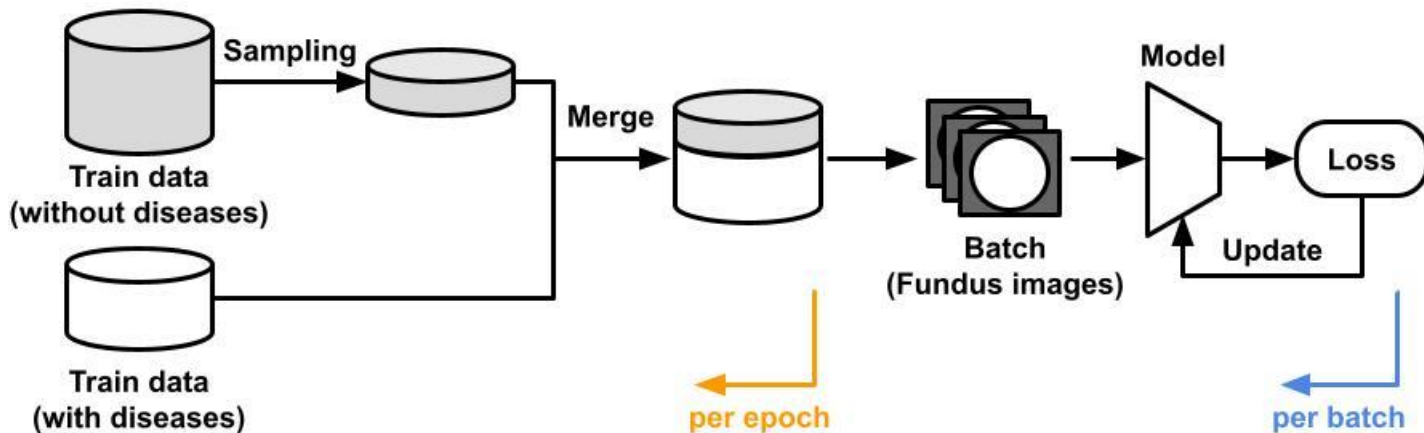
※ Scheie分類

# 方法

- 訓練データ・検証データ・評価データへの分割
  - 疾患が含まれる画像について、各疾患数が8:1:1になるように分割
  - 疾患が含まれない画像については、検証データと評価データについては各500枚に制限し、それ以外を訓練データに入れる

# 方法

- 以上の訓練データを用いて、学習済みのマルチラベル分類モデルを用意した。
- モデルの学習方法



# 方法

- モデルのロバストネス特性を評価する
  - 眼底画像に外乱を疑似的に加える
    - 眼底画像に画像処理を掛ける

# 疑似的な外乱の例



Original



Contrast Up



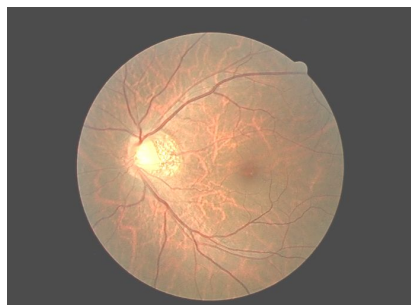
Contrast Down



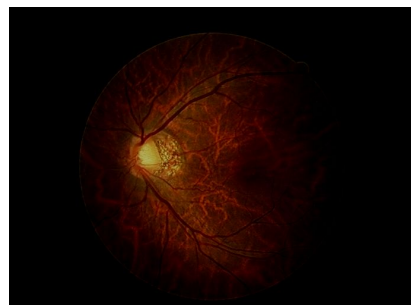
Optical Distortion



Zoom Blur



Brightness Up



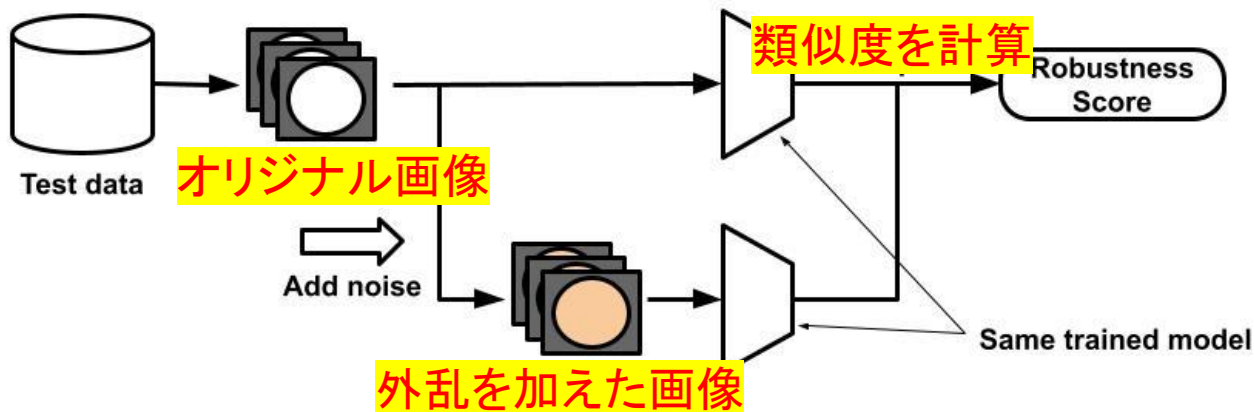
Brightness Down



Zoom In

# 方法

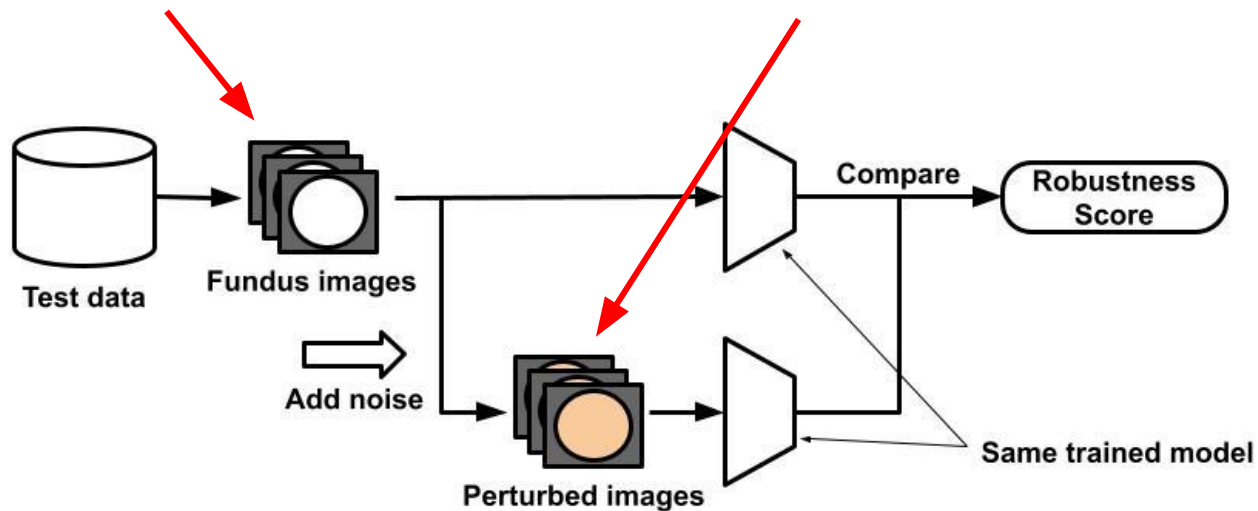
- モデルのロバストネス特性の評価
  - Robustness Scoreを導入する



# 方法

- Robustness Scoreの導入

- オリジナル画像とそれに外乱を加えた画像をモデルに入力

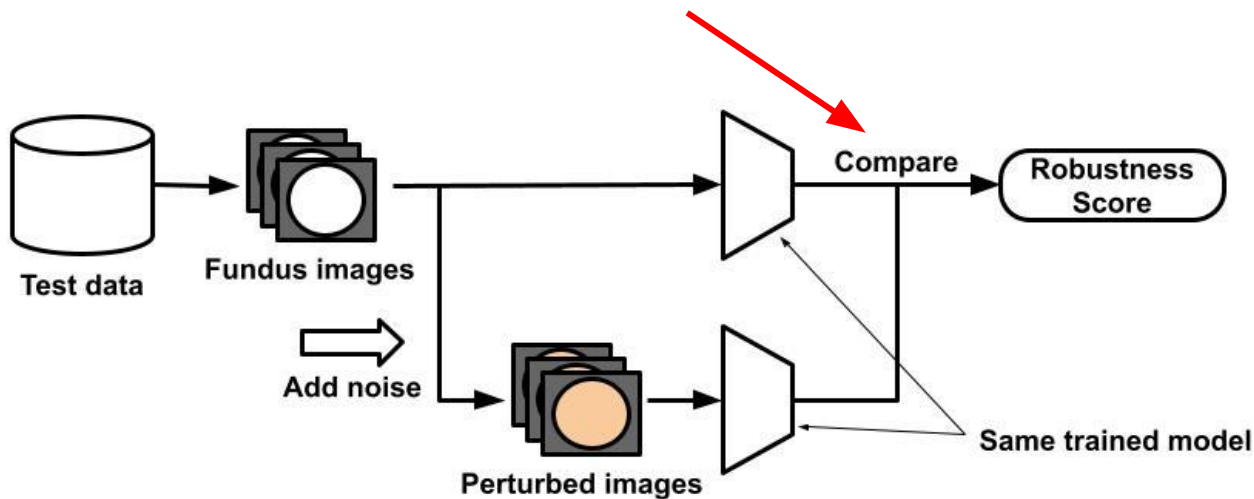




# 方法

- Robustness Scoreの導入

- それぞれの結果の類似度を計算する→Jaccard係数



# 方法

- Robustness Scoreの導入
  - Jaccard係数とは2つの集合の類似度を測る指標
  - $A$ と $B$ を集合として、Jaccard係数 $J(A, B)$ は、

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

# 方法

- Robustness Scoreの導入

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- Jaccard係数の例

- 予測A：オリジナル画像 {H1, S1, 黄斑変性, 眼底出血}
- 予測B：外乱を足した画像 {H2, S1, 黄斑変性}
- 予測A、BのJaccard係数は、

$$J(A, B) = \frac{|S1, 黄斑変性|}{|H1, H2, S1, 黄斑変性, 眼底出血|} = \frac{2}{5} = 0.4$$

# 方法

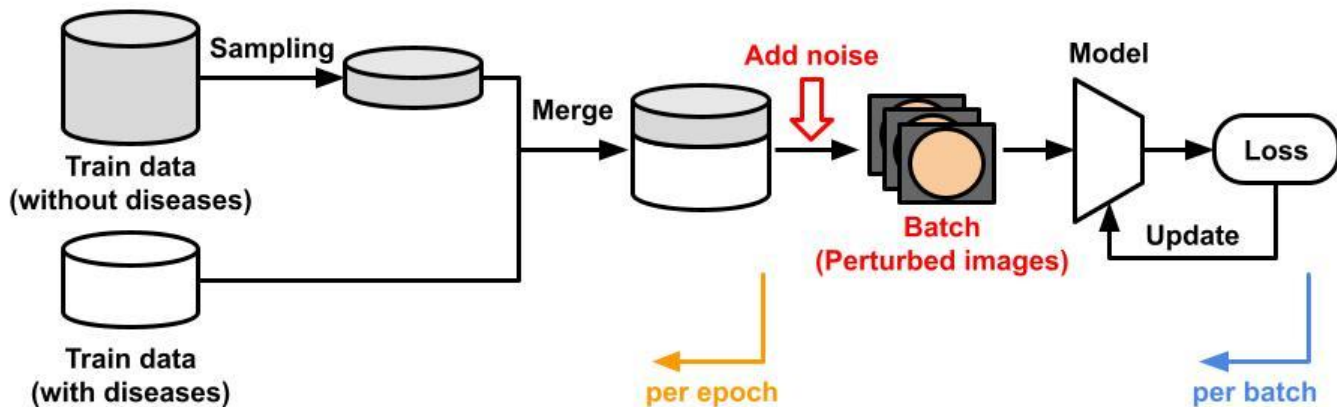
- Robustness Scoreの導入
  - オリジナル画像に対する予測と、外乱を加えた画像に対する予測のJaccard係数を各画像に対して求め、データセット全体で平均を取ったものをRobustness Scoreとする。

$$\text{Robustness Score} = \frac{1}{N} \sum_n J(A_n, B_n)$$

- この指標は0%以上100%以下で、大きくなるほど良い

# 方法

- 様々な外乱についてRobustness Scoreを算出する
- ロバストネス特性の低い外乱を各眼底画像に加えて学習することで、モデルがその外乱への頑強性を付けることができるかを確認する。



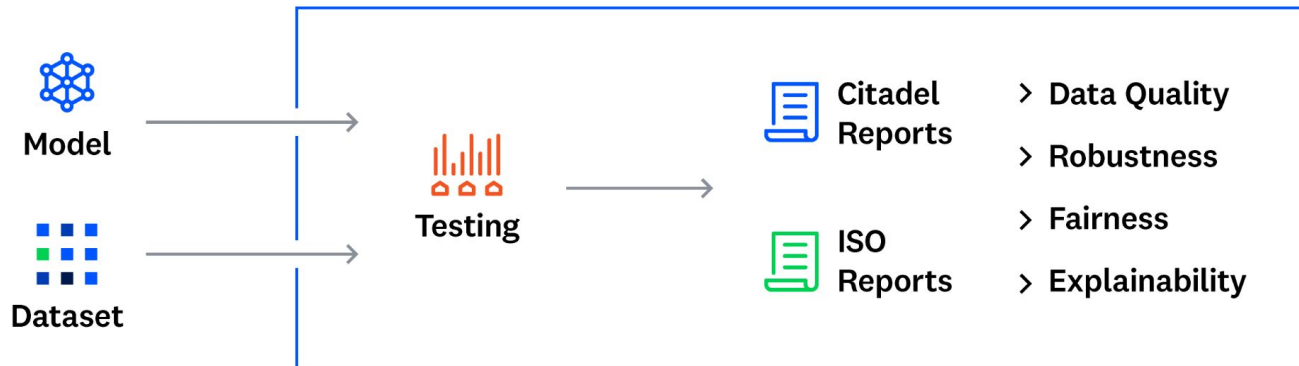
# 方法

- Robustness Scoreの算出とモデル間の性能の比較には、画像分類AIの評価ツールであるCitadel Lensを用いた。

# Citadel Lens

- Citadel AIの提供するソフトウェア
- AIモデルを自動検証し、品質向上を実現

## Citadel Lens



# Citadel Lens

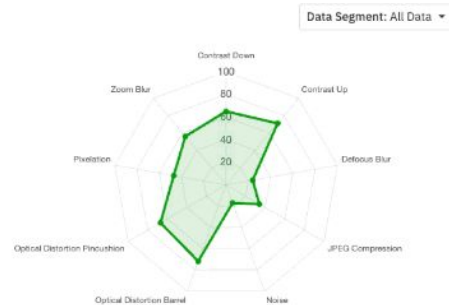
- 網羅的テストを自動かつ一瞬で
  - ノイズ耐性テスト
  - 未学習領域の自動検出
  - ラベル間違い推定
  - 公平性テスト
  - 説明責任の可視化 etc.

## Robustness Scores ②

Camera Robustness 50%

Environment Robustness 60%

Position Robustness 67%

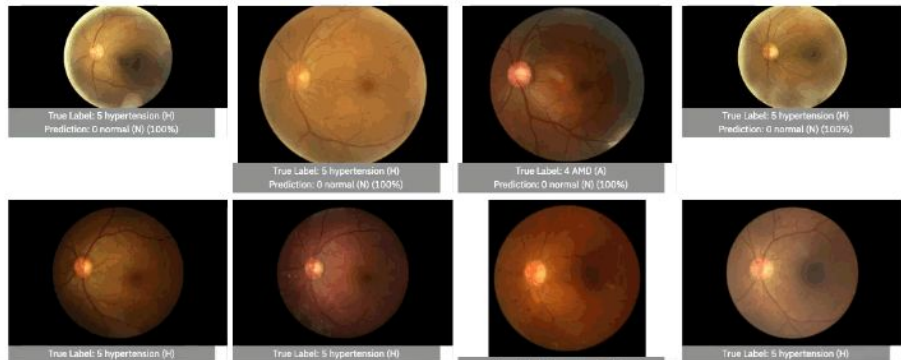


## Image Distortions

Distortion: Camera Robustness - Contrast Down

Sort by: Highest Error

Data Segment: All Data





# Citadel Lens

- 法制度や国際標準への適合性検証とその裏付けとなる技術検証を同時に実現
- 様々なモデルフォーマットに対応

ISO Standards About us News **Taking part** Store

← ISO/IEC JTC 1

## ISO/IEC JTC 1/SC 42

### Artificial intelligence

**25** Published ISO standards +

**31** ISO standards under development +

**39** Participating members



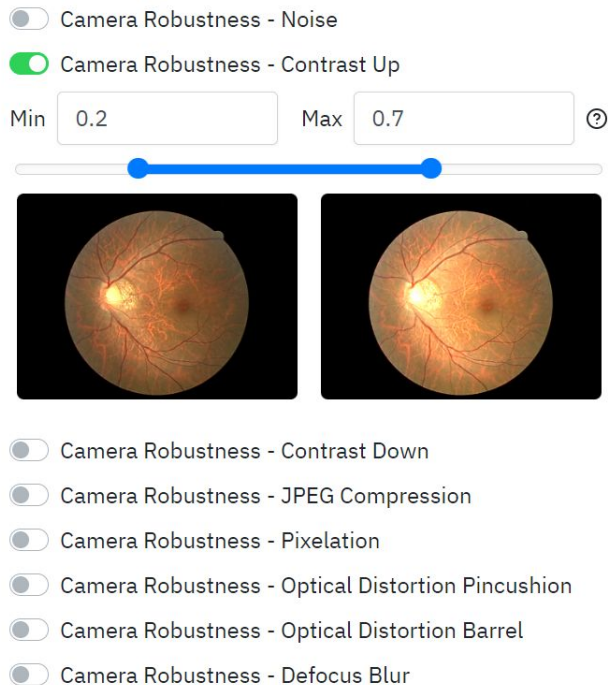
PROJECT NAME	PREDICTION TYPE	DATA TYPE	PROJECT INFO	LATEST REPORT
Chest X-Ray Detection	Object Detection	Image	2 datasets, 2 models, 2 reports	Completed
Traffic Light Object Detection	Object Detection	Image	1 dataset, 1 model, 4 reports	Completed
Credit Risk Prediction	Binary Classification	Tabular	2 datasets, 2 models, 7 reports	Completed
Pathology Patch Classification	Binary Classification	Image	1 dataset, 1 model, 3 reports	Completed
Blood Cell Detection	Object Detection	Image	1 dataset, 1 model, 1 report	Completed
Hospital Readmission Prediction	Binary Classification	Tabular	2 datasets, 2 models, 4 reports	Completed
COCO Object Detection	Object Detection	Image	4 datasets, 2 models, 6 reports	Completed
Hands Gender Classification	Binary Classification	Image	2 datasets, 1 model, 1 report	Completed
Ophthalmic Disease Classification	Multi-Class Classification	Image	1 dataset, 2 models, 5 reports	Completed

# 方法

- Citadel Lensによるロバストネス特性の評価手順

1. 評価したいモデルとデータセット  
をアップロード

2. 一名の医師の元、外乱の種類と  
強度を設定

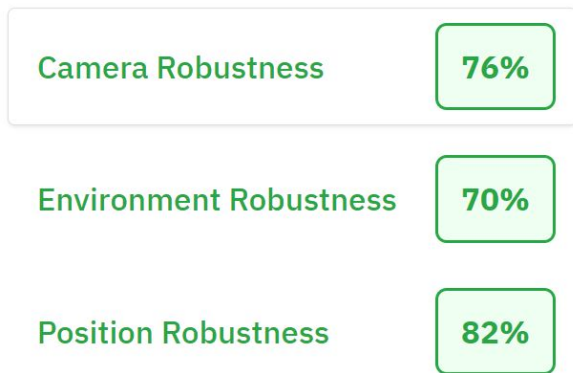


# 方法

- Citadel Lensによるロバストネス特性の評価手順  
3. レポート生成後、各外乱のRobustness Scoreが確認できる。

## Robustness Scores

Data Segment: All Data ▾



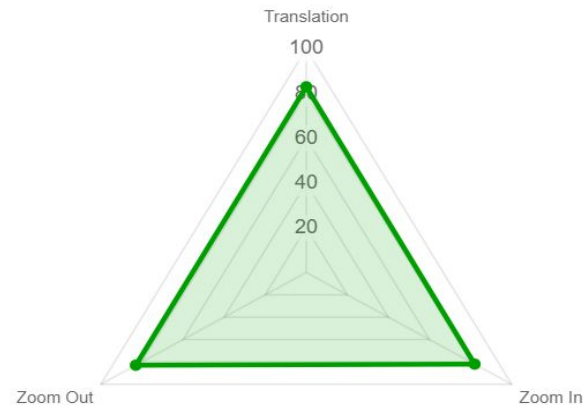
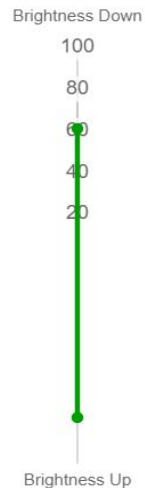
# 評価結果

- Brightness Downの項目でRobustness Scoreが低いことが分かった。
  - 眼底画像の推論は、明度の低さに対する影響を強く受ける。



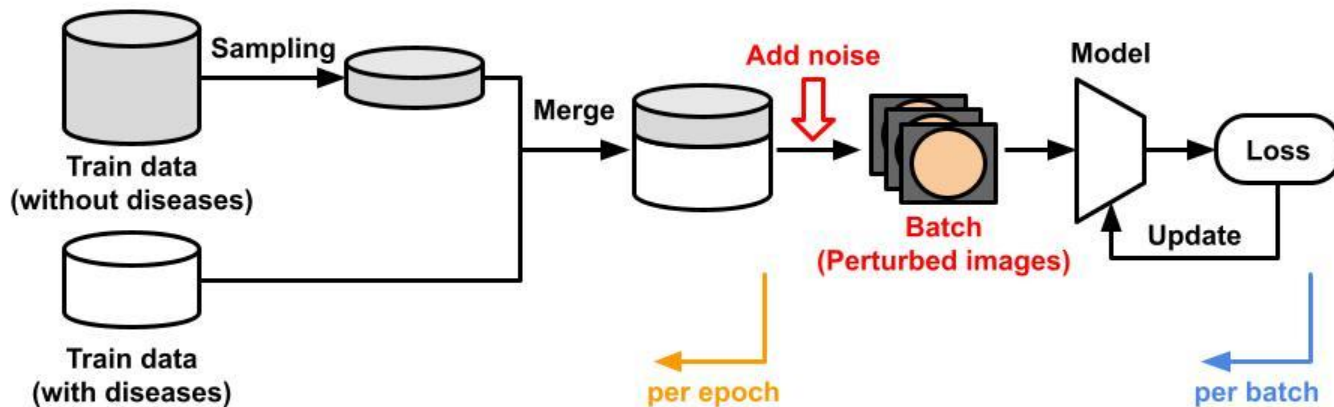
# 評価結果

- **Brightness Down**の項目でRobustness Scoreが低いことが分かった。
  - 眼底画像の推論は、明度の低さに対する影響を強く受ける。



# 改良施策

- 眼底画像への外乱として、眼底画像の明度をランダムに変える処理を追加する。



# 改良施策

- 画像内の各画素に定数を加えて明度を変える。
- 定数の範囲を適切に設定し、画像ごとにランダムな値を加える。

定数の値と画像の変化について



負

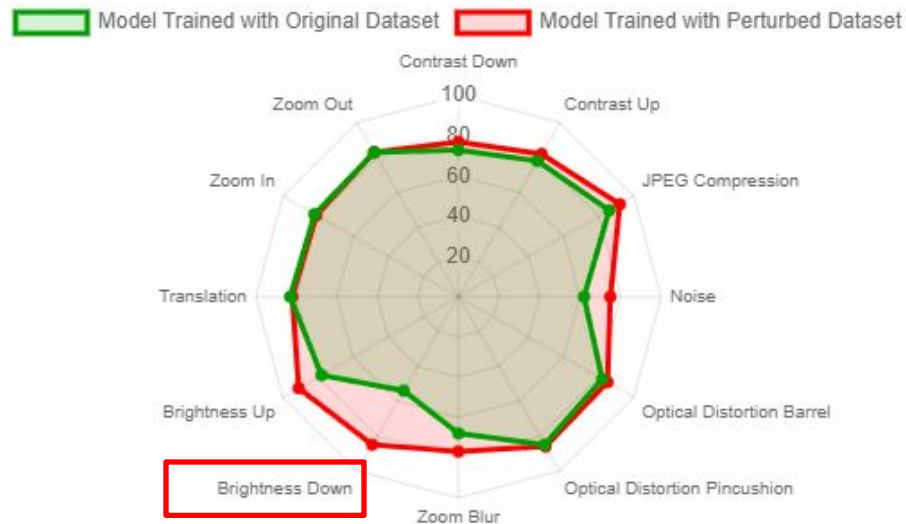
0

正

# 結果

- 明度を変化させて学習させた場合の方が、明度の変化に対するロバストネス特性が高くなることが確認できた。

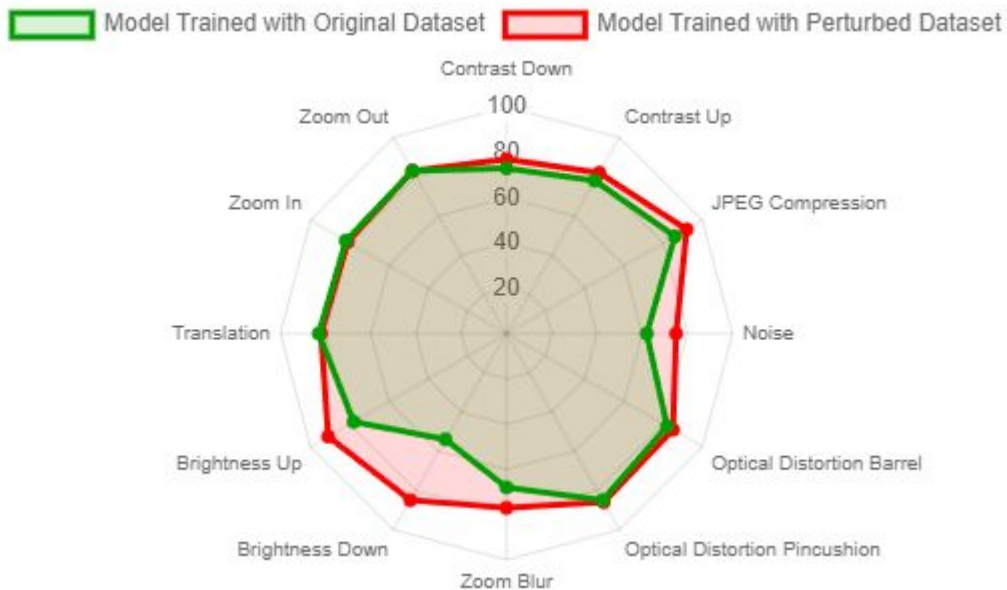
	Brightness Down	Brightness Up
改良前	61%	78%
改良後	<b>89%</b>	<b>91%</b>





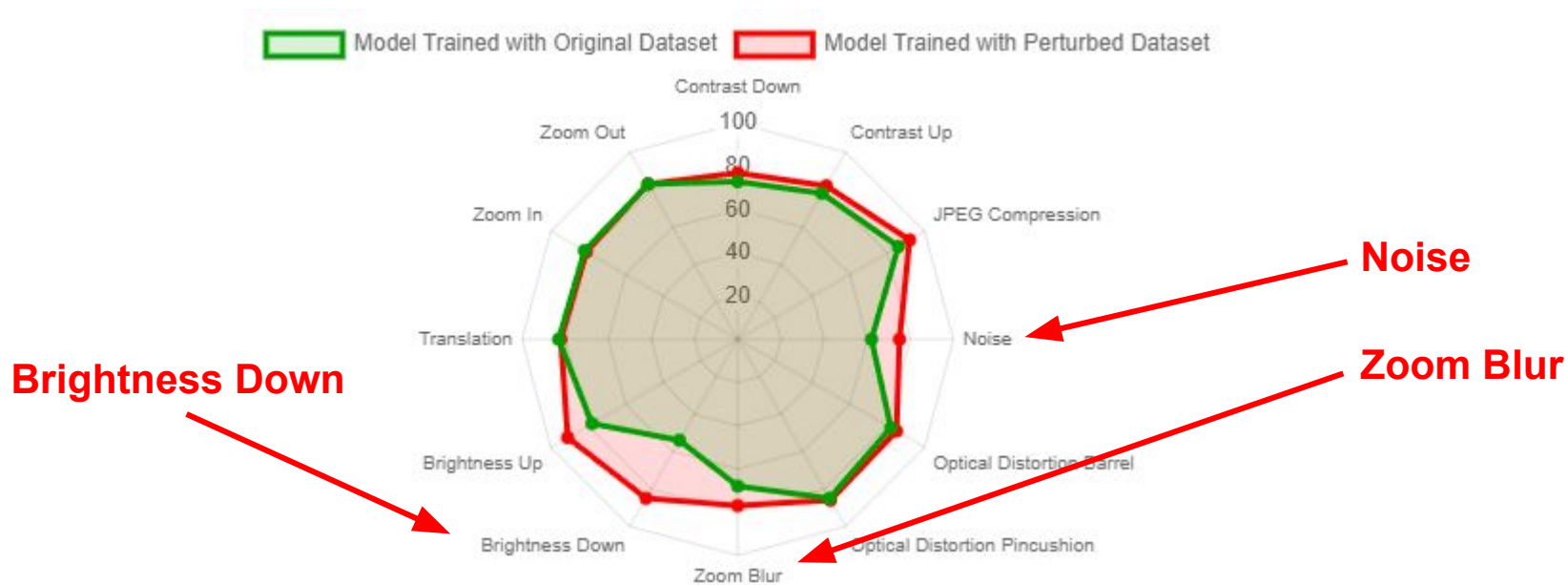
# 考察

- 他の外乱への頑強性も身に付けている。
- 特に、Zoom Blur(ズームぼかし)やNoise(ガウス雑音)



# 考察

- 他の外乱への頑強性も身に付けている。



# 考察

- この二つの外乱に注目する。
- 明度に関して5つの区間に、ただし含まれる画像の枚数が同じになるように分割する。

# 考察

- 元々、両者ともに明度が低い場合にRobustness Scoreが低い。

## ・Zoom Blur

明度	[0.01~0.12]	[0.12~0.15]	[0.15~0.22]	[0.22~0.29]	[0.29~0.7]
改良前	62%	65%	71%	73%	72%
改良後	<b>79%</b>	<b>81%</b>	<b>75%</b>	<b>77%</b>	<b>75%</b>

## ・Noise

明度	[0.01~0.12]	[0.12~0.15]	[0.15~0.22]	[0.22~0.29]	[0.29~0.7]
改良前	39%	46%	57%	82%	<b>85%</b>
改良後	<b>66%</b>	<b>68%</b>	<b>71%</b>	<b>85%</b>	84%

# 考察

- 明度が低い画像についても学習が進み、副次的に他の種類の外乱についても頑強性を得たと思われる。

## ・Zoom Blur

明度	[0.01~0.12]	[0.12~0.15]	[0.15~0.22]	[0.22~0.29]	[0.29~0.7]
改良前	62%	65%	71%	73%	72%
改良後	<b>79%</b>	<b>81%</b>	<b>75%</b>	<b>77%</b>	<b>75%</b>

## ・Noise

明度	[0.01~0.12]	[0.12~0.15]	[0.15~0.22]	[0.22~0.29]	[0.29~0.7]
改良前	39%	46%	57%	82%	<b>85%</b>
改良後	<b>66%</b>	<b>68%</b>	<b>71%</b>	<b>85%</b>	84%

# 結論

- 我々の分類モデルが明度に影響を受けやすいことを定量的に確認した。
- 明度に関する外乱を加えて学習させることで、モデルが明度の変化に対して耐性を得たことを確認した。
- 明度以外の外乱についても同様に学習時に加え、モデルがどのように変化するかを評価するのが今後の課題である。