

要旨

目的

本研究の目的は、複数の医療画像データセットに対し、DA (Data Augmentation) を用いない深層学習 (non-DA) と 3 つの DA 手法、1) RA (Rand Augment) , 2) mixup, 3) SuperMix を用いた深層学習の比較を行うことで、SuperMix の有効性を検証することである。

手法

本研究では、4 つの異なるモダリティから成るオープンデータセットを利用した。各データセットの内訳は、SARS-CoV-2 CT-scan dataset (SARS - CoV-2 の画像 250 枚, SARS-CoV-2 以外の肺疾患の画像 245 枚) , BraTS2019 (脳腫瘍がある画像 300 枚, 脳腫瘍がない画像 299 枚) , DatasetBUSI (良性腫瘍 487 枚, 悪性腫瘍 210 枚) , CMMD (乳癌の画像 274 枚, 非乳癌の画像 276 枚) である。データセットを学習/検証/テストに 60%/20%/20%の比率で分割した。全てのデータセットに対して、DA を用いない深層学習と 3 つの DA 手法、1) RA, 2) mixup, 3) SuperMix を用いた深層学習を行った。テストデータにおける Accuracy, Sensitivity, Specificity, PPV (Positive Predictive Value) , NPV (Negative Predictive Value) , AUC (area under receiver operating

characteristic curve) の 10 回の平均により深層学習モデルを評価した。統計的有意差検定には Delong 検定を用いた。

結果

SARS-CoV-2 CT-scan dataset の各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Average Accuracy: 0.803/0.860/0.759/0.835, Average Sensitivity: 0.740/0.778/0.574/0.804, Average Specificity: 0.867/0.943/0.947/0.867, Average PPV: 0.854/0.933/0.922/0.864, Average NPV: 0.767/0.807/0.691/0.816, Average AUC: 0.872/0.929/0.886/0.924 であった。

BraTS2019 の各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Average Accuracy: 0.798/0.826/0.773/0.843, Average Sensitivity: 0.737/0.765/0.617/0.773, Average Specificity: 0.858/0.887/0.928/0.913, Average PPV: 0.845/0.870/0.899/0.901, Average NPV: 0.779/0.795/0.710/0.803, Average AUC: 0.887/0.917/0.905/0.936 であった。

DatasetBUSI の各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix

の順に Average Accuracy: 0.678/0.703/0.731/0.667, Average Sensitivity: 0.719/0.800/0.405/0.845, Average Specificity: 0.658/0.656/0.889/0.581, Average PPV: 0.506/0.534/0.649/0.503, Average NPV: 0.832/0.873/0.758/0.892, Average AUC: 0.761/0.799/0.764/0.809 であった。

CMMD の各実験の平均の評価指標は, non-DA/RA/mixup/SuperMix の順に Average Accuracy: 0.672/0.690/0.691/0.635, Average Sensitivity: 0.493/0.493/0.505/0.350, Average Specificity: 0.855/0.891/0.880/0.926, Average PPV: 0.778/0.829/0.843/0.831, Average NPV: 0.629/0.637/0.649/0.588, Average AUC: 0.772/0.811/0.812/0.824 であった。

SuperMix は RA と同等もしくはその他の手法と比べて有意に高い結果になった。

結論

複数の医療画像を用いた分類タスクにおいて, SuperMix による DA は従来の DA 手法と比較して, 分類精度の向上に有効であることが示唆された。