

## INTRODUCTION

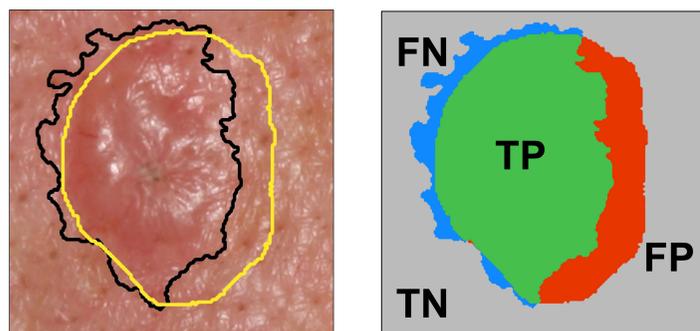
- Skin lesion images usually have a considerable portion of **background healthy skin**.
- Cross entropy-based loss values are a poor reflection of the segmentation quality, thus **overlap-based losses are preferred** [1].
- Popular overlap-based segmentation loss functions (Dice and IoU loss) **do not account for true negative pixel predictions**.

## AIM

Propose a differentiable overlap-based loss function which:

- **penalizes misclassification of the background pixels** (true negative predictions).
- tackles class imbalance between the foreground and the background pixels, thus achieving a **better sensitivity-specificity trade-off**.

## METHOD



Matthews correlation coefficient (MCC) has been shown to be an **optimal metric** for skewed distributions and imbalanced classes [2, 3].

MCC values range from  $-1$  (completely disjoint prediction) to  $+1$  (perfect prediction).

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$\mathcal{L}_{MCC} = 1 - MCC$$

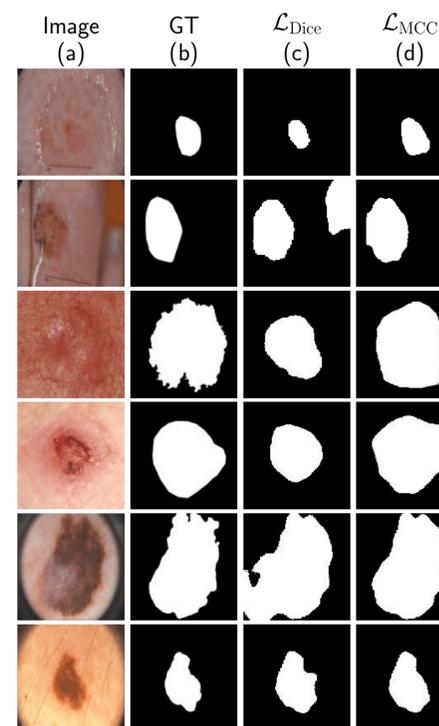
## RESULTS

### Experimental Details

- **U-Net segmentation network** with encoder-decoder architecture and skip connections.
- Models trained with **MCC loss ( $\mathcal{L}_{MCC}$ )** and **Dice loss ( $\mathcal{L}_{Dice}$ )** with **data augmentation** (horizontal and vertical flipping and rotation).
- Evaluated on **3 datasets**:
  - ISBI ISIC 2017 (600 test images)
  - DermoFit (390 test images)
  - PH2 (60 test images)

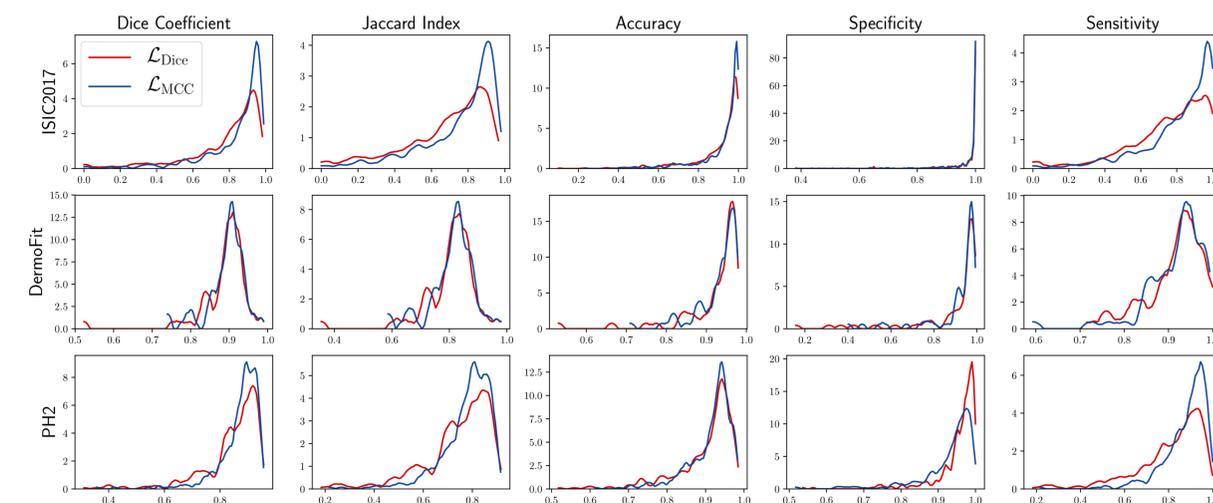
### Key Observations

- Models trained with  $\mathcal{L}_{MCC}$  **yield higher performance metrics** than those trained with  $\mathcal{L}_{Dice}$  for all cases except one, where  $\mathcal{L}_{MCC}$  yields a better balance between sensitivity and specificity.
- KDE plots show higher peaks (i.e., **larger densities**) at **higher values** for all metrics.
- For ISIC 2017,  $\mathcal{L}_{MCC}$  **results are within 1% of the top 3 on the challenge leaderboard** without any post-processing, external data, or model ensembles.
- All improvements in Jaccard index are **statistically significant** (\*\*\*:  $p < 0.001$  and \*:  $p < 0.05$ ).



Qualitative results for skin lesion segmentation (rows 1-2: ISIC 2017, rows 3-4: DermoFit, rows 5-6: PH2). Note the large fraction of background.

Dataset	ISIC 2017***		DermoFit***		PH2*	
Loss Function	$\mathcal{L}_{Dice}$	$\mathcal{L}_{MCC}$	$\mathcal{L}_{Dice}$	$\mathcal{L}_{MCC}$	$\mathcal{L}_{Dice}$	$\mathcal{L}_{MCC}$
Dice	$0.778 \pm 0.007$	<b><math>0.838 \pm 0.007</math></b>	$0.844 \pm 0.004$	<b><math>0.871 \pm 0.003</math></b>	$0.889 \pm 0.002$	<b><math>0.894 \pm 0.002</math></b>
Jaccard	$0.676 \pm 0.010$	<b><math>0.752 \pm 0.008</math></b>	$0.742 \pm 0.005$	<b><math>0.778 \pm 0.004</math></b>	$0.805 \pm 0.004$	<b><math>0.811 \pm 0.003</math></b>
Accuracy	$0.903 \pm 0.005$	<b><math>0.922 \pm 0.005</math></b>	$0.902 \pm 0.006$	<b><math>0.914 \pm 0.002</math></b>	$0.922 \pm 0.003$	<b><math>0.930 \pm 0.002</math></b>
Sensitivity	$0.747 \pm 0.009$	<b><math>0.813 \pm 0.008</math></b>	$0.808 \pm 0.006$	<b><math>0.880 \pm 0.004</math></b>	$0.913 \pm 0.003$	<b><math>0.946 \pm 0.003</math></b>
Specificity	$0.968 \pm 0.003$	<b><math>0.971 \pm 0.003</math></b>	<b><math>0.953 \pm 0.002</math></b>	$0.940 \pm 0.003$	$0.885 \pm 0.007$	<b><math>0.908 \pm 0.005</math></b>



## CONCLUSION

- We propose a **novel metric-based differentiable loss function** for binary segmentation using the **Matthews correlation coefficient**, which incorporates a **penalty for misclassifying the true negative pixels**.
- MCC loss **improves skin lesion segmentation performance** on 3 datasets: ISBI ISIC 2017, DermoFit, and PH2 as compared to Dice loss, and the improvements are **statistically significant**.
- MCC loss achieves a **better trade-off** between sensitivity and specificity.
- Future work involves **generalizing the loss for K classes** and evaluating on **more segmentation model architectures** and **medical imaging modalities**.

## REFERENCES

- [1] Berman M et al. "The Lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks". CVPR 2018.
- [2] MaQC Consortium et al. "The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models". Nat. Biotechnol. 2010.
- [3] Boughorbel S et al. "Optimal classifier for imbalanced data using Matthews correlation coefficient metric". PloS One 2017.

## ACKNOWLEDGEMENTS

We thank Natural Sciences and Engineering Research Council of Canada (**NSERC**) and Canadian Institutes of Health Research (**CIHR**) for funding, and **NVIDIA Corporation** for compute resources.

## CONTACT INFORMATION

Kumar Abhishek, PhD student, SFU ([kabhish@sfu.ca](mailto:kabhish@sfu.ca))

[www.MedicalImageAnalysis.com](http://www.MedicalImageAnalysis.com)

