Automated malignant melanoma detection using supervised contrastive learning

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Cassandra Durr
Supervisor: Dr. Hanno Coetzer

Photo by Stefan Els

Introduction

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together sonke siya phambili saam vorentoe

This study investigates supervised contrastive learning and diverse encoder architectures for improving melanoma detection classification results.

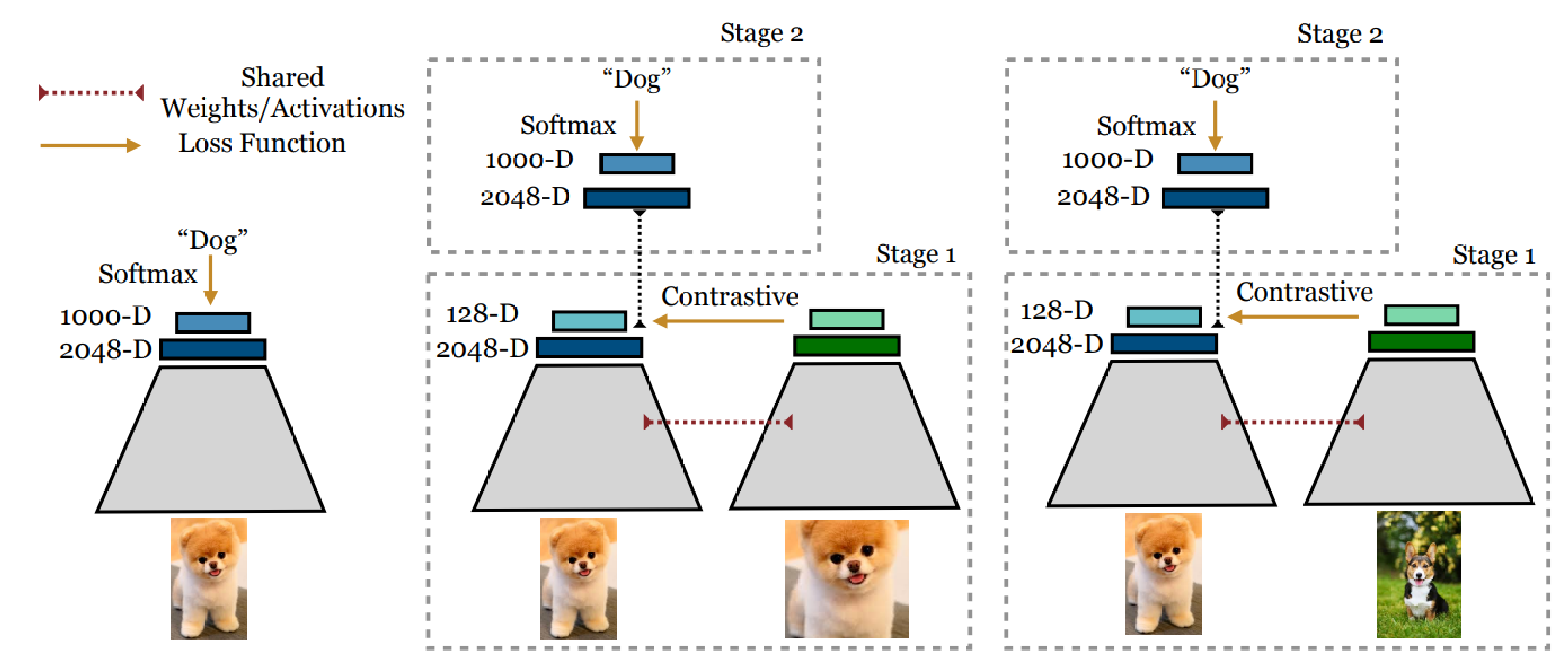
- Challenges associated with melanoma detection
- Promise of computer-aided melanoma detection
- Favourable ImageNet classification results from ‘Supervised Contrastive Learning’ paper by Khosla et al. (2020).
- Encoder architectures explored:
  - Vision transformer
  - ResNet50
  - InceptionV3

Problem statement

1. Does supervised contrastive learning outperform traditional image classification models in detecting melanoma?
   - Hypothesis: Supervised contrastive learning has shown great promise in literature, producing state-of-the-art classification performance.
2. Do vision transformers yield superior classification performance over CNNs for the task of melanoma detection?
   - Hypothesis : Vision transformers have an enhanced long-range spatial awareness, resulting in impressive performance in recent literature.

Baseline and Supervised Contrastive Learning Methodology

Supervised Cross Entropy vs Self-supervised CL vs Supervised CL



(a) Supervised Cross Entropy (b) Self Supervised Contrastive (c) Supervised Contrastive

Ref:
Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. 2020.

Supervised contrastive learning & the SupCon Loss

Supervised contrastive loss pushes encodings from the same class closer together in the embedding space while pulling apart encodings from different classes.

Supervised contrastive loss function, SupCon:

Losses per image in the batch

Large dot product when encodings are well-aligned

$$L = \sum_{i \in I} L_i$$

$$= \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \ln \left(\frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_a / \tau)} \right)$$

Small values when negatives are well-separated.

- z_i: encoding of the anchor image
- z_p: encoding of an image from the same class
- z_a: encoding of an image with a different class label

Supervised contrastive learning & the SupCon Loss

Supervised contrastive loss pushes encodings from the same class closer together in the embedding space while pulling apart encodings from different classes.

Supervised contrastive loss function, SupCon:

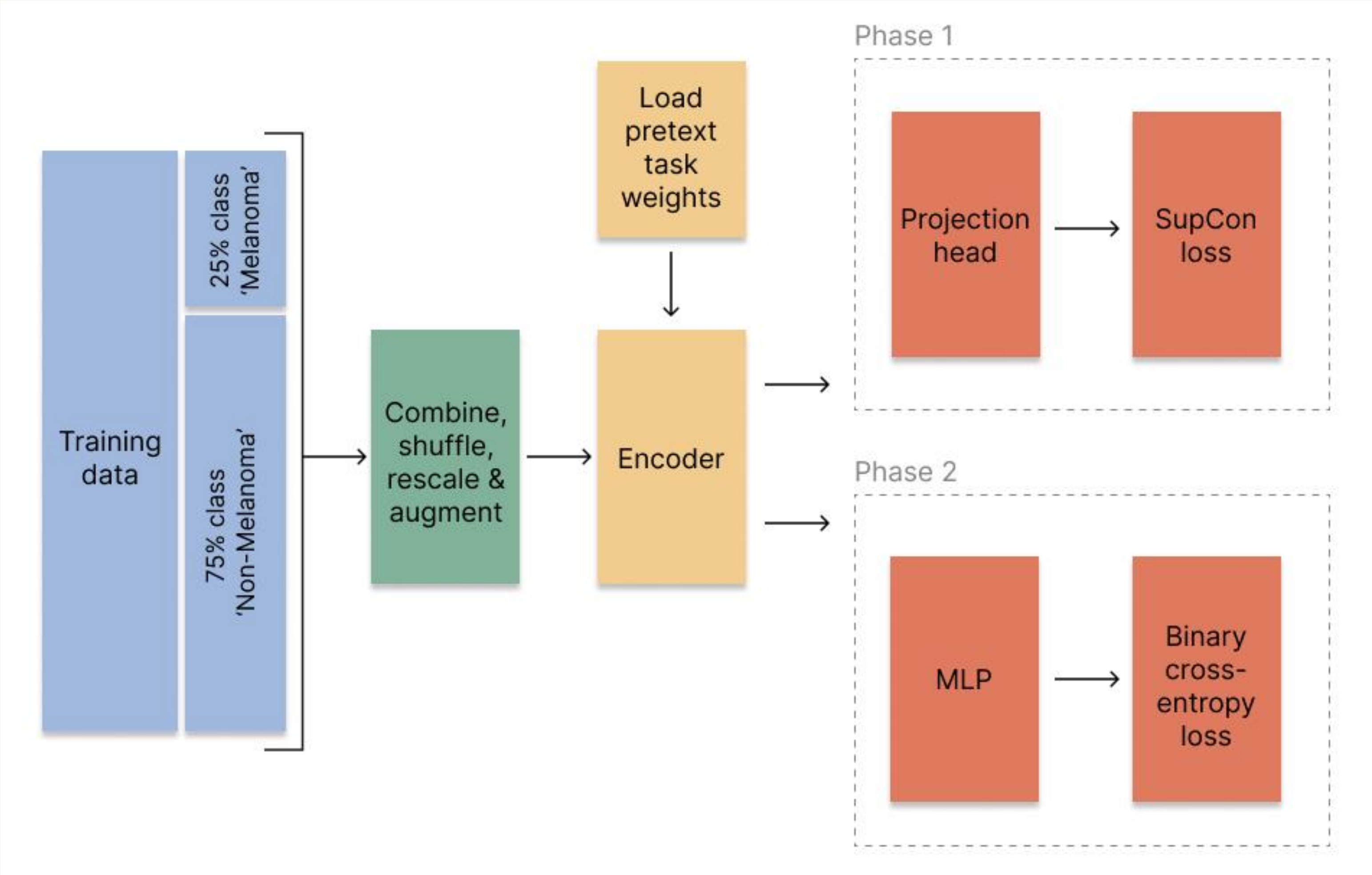
Negate increasingly large values – small loss values, shrinking as representation space improves.

$$
\begin{aligned}
L &= \sum_{i \in I} L_i \\
&= \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \ln \left(\frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_a / \tau)} \right)
\end{aligned}
$$

Increasingly large value divided by shrinking values
> Increasingly large values

Methodology

- Class-weighted batching scheme
- Data processing & augmentation
- Encoding
- Pretext training
- Staged training method
- F_2 prediction threshold



Validation Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together sonke siya phambili saam vorentoe

Summary

1. SupCon vs Baseline across encoders
2. Small vs larger batch size
3. Relative SupCon performance as batch size increases

Encoder		ViT		ResNet50		InceptionV3	
Encoder parameters		34,276,288		23,564,800		21,802,784	
Training method		Base	SupCon	Base	SupCon	Base	SupCon
Batch size: 16	Threshold	0.2455	0.2859	0.2131	0.2212	0.2212	0.2455
	AUC	0.8829	0.8759	0.8589	0.8889	0.8792	0.8740
	Accuracy	0.7109	0.7482	0.7728	0.7655	0.7568	0.6969
	Recall	0.9277	0.8735	0.8072	0.8916	0.8735	0.9398
	F_2	0.6210	0.6218	0.6052	0.6457	0.6288	0.6166
	Best score	2	2	1	3	3	1
	Mean score	0.7856	0.7798	0.7610	0.7979	0.7846	0.7818
Batch size: 64	Threshold	0.2939	0.1646	0.2455	0.2374	0.1889	0.3101
	AUC	0.8646	0.7604	0.8681	0.8750	0.8703	0.8741
	Accuracy	0.7295	0.5396	0.7688	0.6969	0.7022	0.7722
	Recall	0.8675	0.9458	0.7952	0.9277	0.9036	0.8373
	F_2	0.6040	0.5223	0.5951	0.6106	0.6024	0.6216
	Best score	3	1	1	3	1	3
	Mean score	0.7664	0.6920	0.7568	0.7776	0.7696	0.7763

*SupCon = supervised contrastive learning

Validation Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

forward together sonke siya phambili saam vorentoe

Summary

1. SupCon vs Baseline across encoders
2. Small vs larger batch size
3. Relative SupCon performance as batch size increases

Encoder		ViT		ResNet50		InceptionV3	
Encoder parameters		34,276,288		23,564,800		21,802,784	
Training method		Base	SupCon	Base	SupCon	Base	SupCon
Batch size: 16	Threshold	0.2455	0.2859	0.2131	0.2212	0.2212	0.2455
	AUC	0.8829	0.8759	0.8589	0.8889	0.8792	0.8740
	Accuracy	0.7109	0.7482	0.7728	0.7655	0.7568	0.6969
	Recall	0.9277	0.8735	0.8072	0.8916	0.8735	0.9398
	F_2	0.6210	0.6218	0.6052	0.6457	0.6288	0.6166
	Best score	2	2	1	3	3	1
	Mean score	0.7856	0.7798	0.7610	0.7979	0.7846	0.7818
Batch size: 64	Threshold	0.2939	0.1646	0.2455	0.2374	0.1889	0.3101
	AUC	0.8646	0.7604	0.8681	0.8750	0.8703	0.8741
	Accuracy	0.7295	0.5396	0.7688	0.6969	0.7022	0.7722
	Recall	0.8675	0.9458	0.7952	0.9277	0.9036	0.8373
	F_2	0.6040	0.5223	0.5951	0.6106	0.6024	0.6216
	Best score	3	1	1	3	1	3
	Mean score	0.7664	0.6920	0.7568	0.7776	0.7696	0.7763

*SupCon = supervised contrastive learning

Validation Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together sonke siya phambili saam vorentoe

Summary

1. SupCon vs Baseline across encoders
2. Small vs larger batch size
3. Relative SupCon performance as batch size increases

Encoder		ViT		ResNet50		InceptionV3	
Encoder parameters		34,276,288		23,564,800		21,802,784	
Training method		Base	SupCon	Base	SupCon	Base	SupCon
Batch size: 16	Threshold	0.2455	0.2859	0.2131	0.2212	0.2212	0.2455
	AUC	0.8829	0.8759	0.8589	0.8889	0.8792	0.8740
	Accuracy	0.7109	0.7482	0.7728	0.7655	0.7568	0.6969
	Recall	0.9277	0.8735	0.8072	0.8916	0.8735	0.9398
	F_2	0.6210	0.6218	0.6052	0.6457	0.6288	0.6166
	Best score	2	2	1	3	3	1
	Mean score	0.7856	0.7798	0.7610	0.7979	0.7846	0.7818
Batch size: 64	Threshold	0.2939	0.1646	0.2455	0.2374	0.1889	0.3101
	AUC	0.8646	0.7604	0.8681	0.8750	0.8703	0.8741
	Accuracy	0.7295	0.5396	0.7688	0.6969	0.7022	0.7722
	Recall	0.8675	0.9458	0.7952	0.9277	0.9036	0.8373
	F_2	0.6040	0.5223	0.5951	0.6106	0.6024	0.6216
	Best score	3	1	1	3	1	3
	Mean score	0.7664	0.6920	0.7568	0.7776	0.7696	0.7763

*SupCon = supervised contrastive learning

Validation Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT | forward together sonke siya phambili saam vorentoe

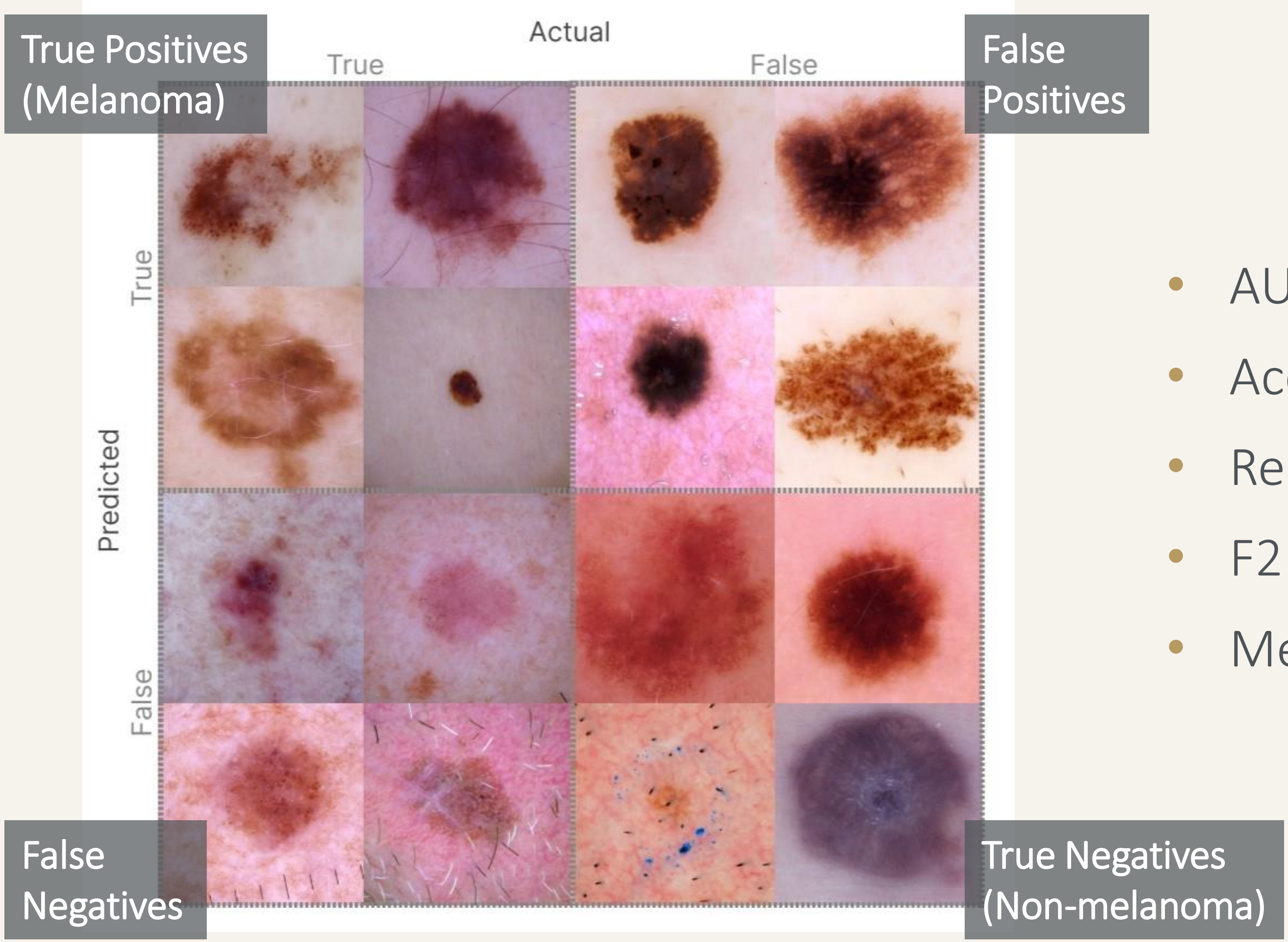
Summary

1. SupCon vs Baseline across encoders
2. Small vs larger batch size
3. Relative SupCon performance as batch size increases

Encoder		ViT		ResNet50		InceptionV3	
Encoder parameters		34,276,288		23,564,800		21,802,784	
Training method		Base	SupCon	Base	SupCon	Base	SupCon
Batch size: 16	Threshold	0.2455	0.2859	0.2131	0.2212	0.2212	0.2455
	AUC	0.8829	0.8759	0.8589	0.8889	0.8792	0.8740
	Accuracy	0.7109	0.7482	0.7728	0.7655	0.7568	0.6969
	Recall	0.9277	0.8735	0.8072	0.8916	0.8735	0.9398
	F_2	0.6210	0.6218	0.6052	0.6457	0.6288	0.6166
	Best score	2	2	1	3	3	1
	Mean score	0.7856	0.7798	0.7610	0.7979	0.7846	0.7818
Batch size: 64	Threshold	0.2939	0.1646	0.2455	0.2374	0.1889	0.3101
	AUC	0.8646	0.7604	0.8681	0.8750	0.8703	0.8741
	Accuracy	0.7295	0.5396	0.7688	0.6969	0.7022	0.7722
	Recall	0.8675	0.9458	0.7952	0.9277	0.9036	0.8373
	F_2	0.6040	0.5223	0.5951	0.6106	0.6024	0.6216
	Best score	3	1	1	3	1	3
	Mean score	0.7664	0.6920	0.7568	0.7776	0.7696	0.7763

*SupCon = supervised contrastive learning

Test Results



- AUC: 0.8569
- Accuracy: 0.7055
- Recall: 0.8304
- F2: 0.5717
- Mean score: 0.7411

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Problem statement

1. Does supervised contrastive learning outperform traditional image classification models in detecting melanoma?
   - Finding: Supervised contrastive learning does **not** consistently produce superior melanoma detection performance.
2. Do vision transformers yield superior classification performance over CNNs for the task of melanoma detection?
   - Finding : ViT encoders do **not** necessarily outperform CNN architectures.

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Thank you
Enkosi
Dankie

Photo by Stefan Els