# Impact of Quality of Image Database on AI Performance in Skin Cancer Detection

Yixuan Li

Supervisor: Prof. Patrice Delmas

22 November 2022

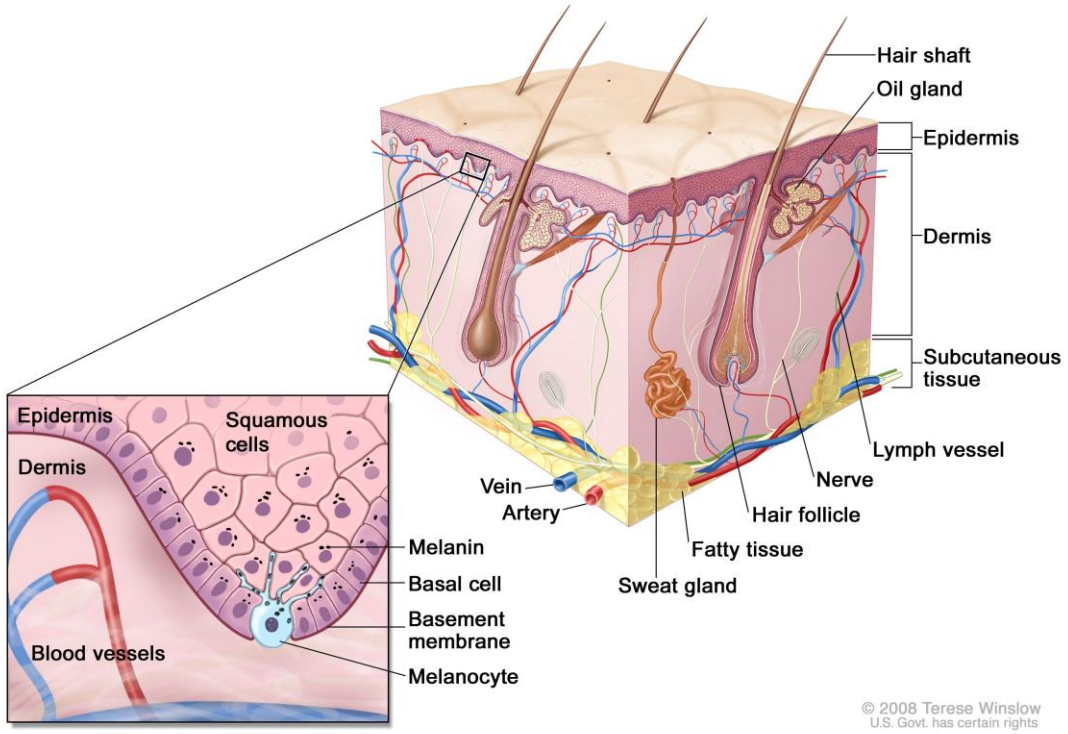
# Motivation

- > 82,000 New Zealanders get skin cancer every year
- Due to depletion of ozone layer, this number is increasing 10-20% annually, accounting 80% of all new cancers in NZ.
- Melanoma is the most deadly one in all skin cancers, accounting <5% of all skin cancers, but responsible for >75% of total death.
- the highest incidence rate from melanoma NZ > Australia > European countries.
- Most skin cancers can be prevented if detected early.

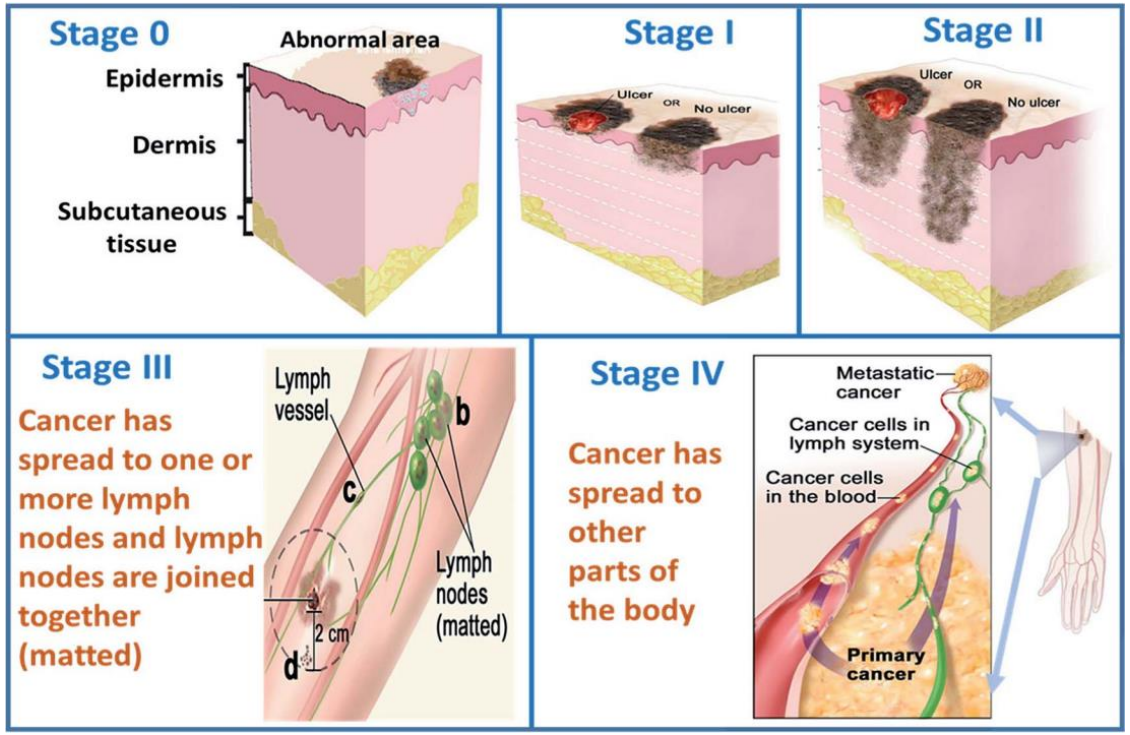
Major cause : Over-exposure of UV rays

# Introduction

- Early detection would increase Melanoma 5 years' survival rate from 15% to 99%.

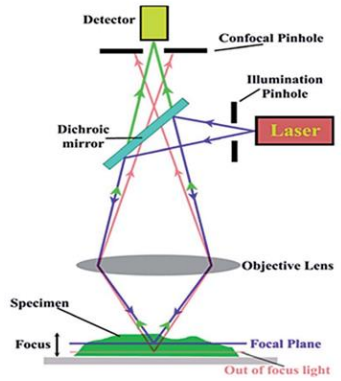



Normal Skin




Five Stages of Skin Cancer

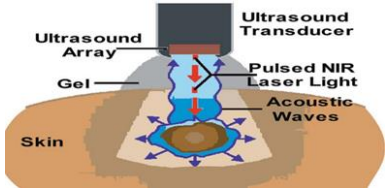
# Non-Invasive Detection of Skin Cancer



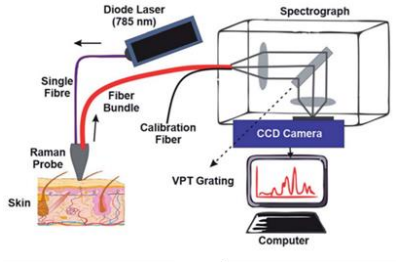

Reflectance Confocal Microscopy

Dermatoscopy

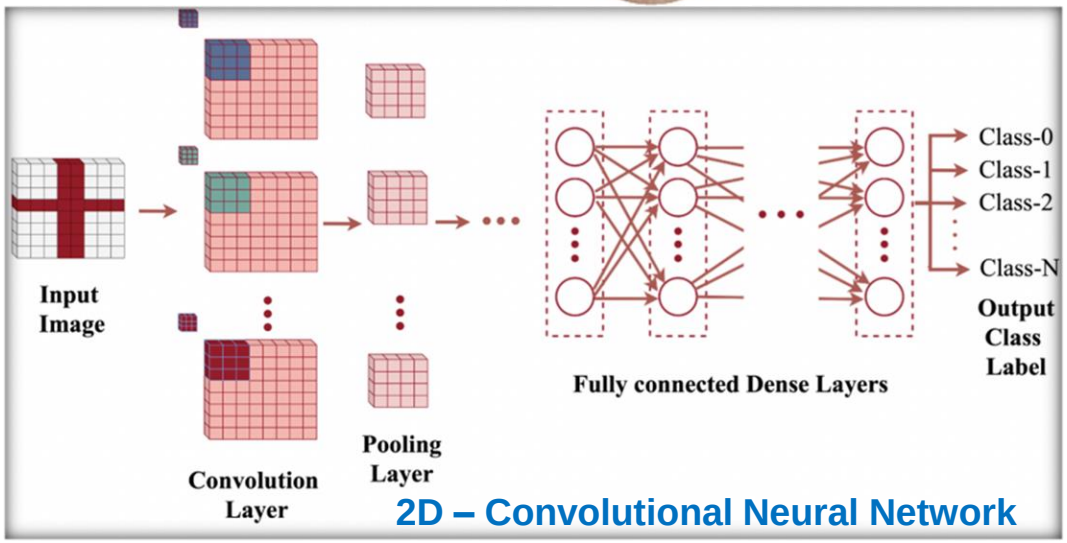
Fast, Easy Accessible, Low-cost,



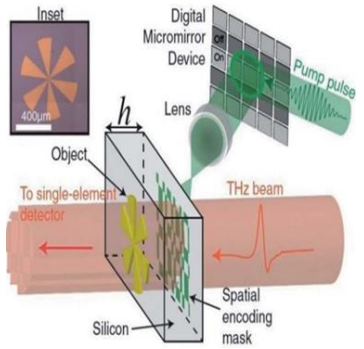

High-frequency ultrasound




Raman spectroscopy




2D – Convolutional Neural Network



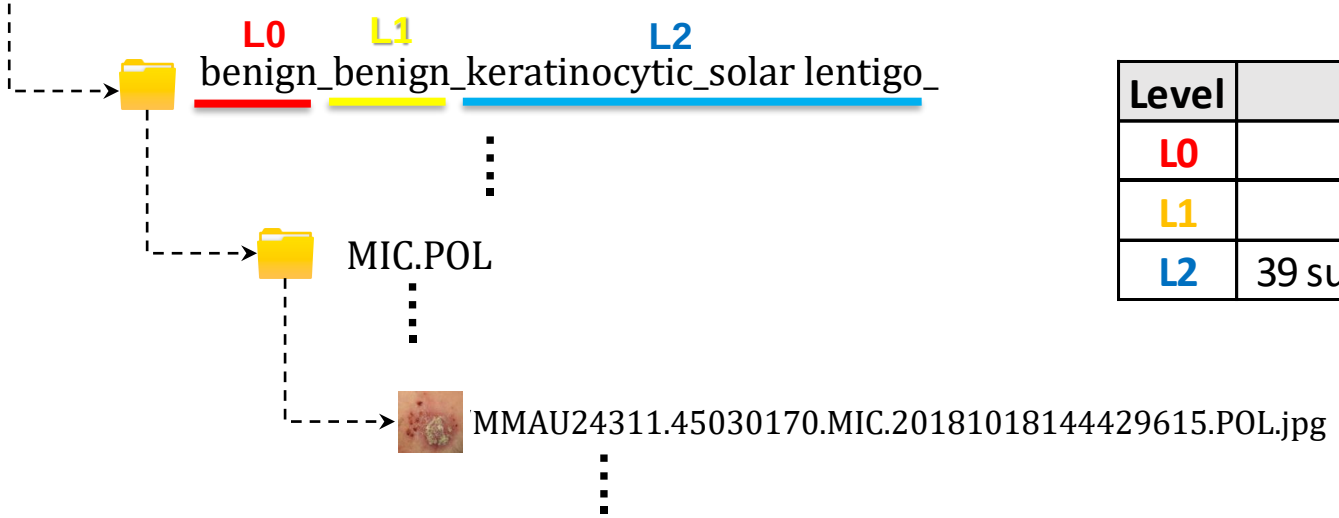

Tetrahertz Spectroscopy

# Research Questions

- How to track, analyze, manage image database modification
- How to track uniqueness of image files and archive them in SQL database
- What is the relationship of the quality of AI Database to the model performance

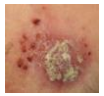
# Methods

- Structure of AI database: three-level labeling system




| Level | Classes |
|---|---|
| L0 | Benign , Malignant |
| L1 | Benign, IEC, Melanoma, NMSC |
| L2 | 39 subclasses, e.g.,Dermatofibroma, Keratinocytic |

- Image-hashing to track uniqueness of images



SHA256 →

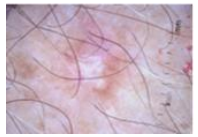
Hash Code (256 bits)

44f3142b5c659e96ca4557b5d45f061e3479ca49aa3e0ed4a27fcb412502d88f

# Statistical Analysis on AI database

Three kinds of errors in AI database:

## 1, 0.15% Redundancy due to multi-upload error



ASP305456.31550590.MIC.20140428080042946.POL.jpg
ASP701991.31550590.MIC.20140428080042946.POL.jpg
ASPH03700.31550590.MIC.20140428080042946.POL.jpg

## 2, 0.22% Cross-labelling

| name | path | hashcode |
|---|---|---|
| @POD00004.14970729.MIC.20140503120707252.POL.jpg | malignant_melanoma_melanoma_/MIC | 30fa4765b3cd19236d59cfef45faf25d01ecd4d15e95c3f41887e997d10bb846 |
| @POD00004.14970729.MIC.20140503120707252.POL.jpg | benign_benign_vascular_telangiectasia_/benign_benign_nevus benign_/MIC.POL | 30fa4765b3cd19236d59cfef45faf25d01ecd4d15e95c3f41887e997d10bb846 |

## 3, 22.56% Repetitive - labelling

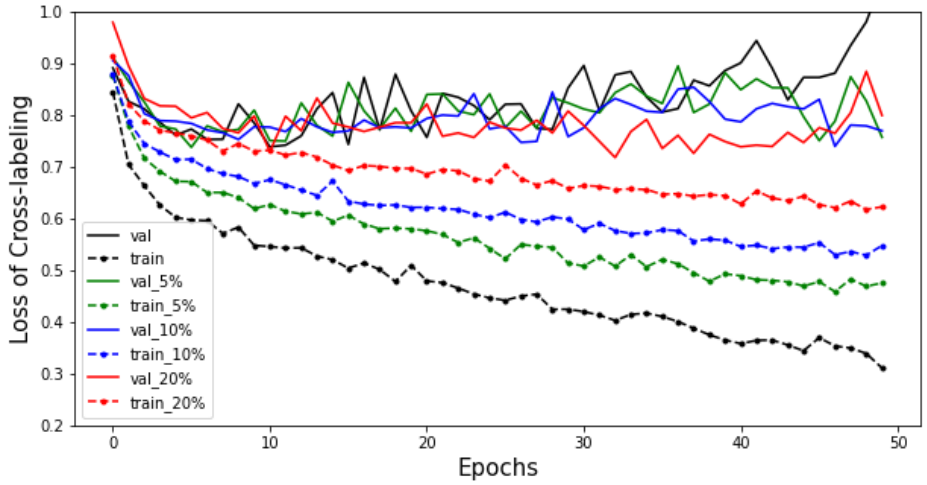
Deleted + some re-labellings

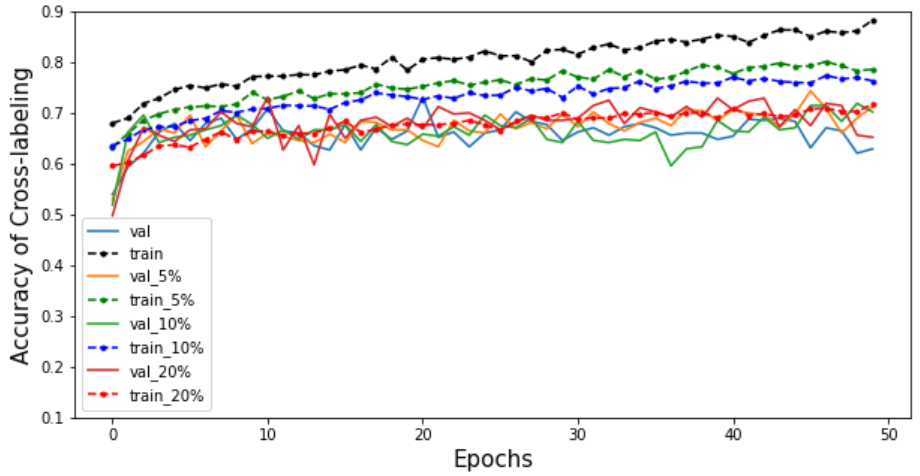
| Derivations | L0 | L1 melanoma | L2 melanoma |
|---|---|---|---|
| Sensitivity(TPR) | 0.95 | 0.95 | 0.95 |
| Specificity(TNR) | 0.72 | 0.69 | 0.69 |
| Precision (PPV) | 0.74 | 0.17 | 0.18 |
| Distribuion of lesion(1/PPV) | 1.36 | 5.77 | 5.70 |
| Negative predictive value NPV | 0.95 | 1.00 | 1.00 |
| pevalence threshold(PT) | 0.35 | 0.37 | 0.36 |
| F1-score | 0.83 | 0.29 | 0.30 |
| Accuracy(ACC) | 0.82 | 0.70 | 0.71 |
| After cleaning | | | |
| Sensitivity(TPR) | 0.95 | 0.95 | 0.95 |
| Specificity(TNR) | 0.72 | 0.69 | 0.70 |
| Precision (PPV) | 0.74 | 0.20 | 0.21 |
| Distribuion of lesion(1/PPV) | 1.36 | 4.93 | 4.86 |
| Negative predictive value NPV | 0.95 | 0.99 | 0.99 |
| pevalence threshold(PT) | 0.35 | 0.36 | 0.36 |
| F1-score | 0.83 | 0.34 | 0.34 |
| Accuracy(ACC) | 0.82 | 0.71 | 0.72 |

# Simulation with Melanoma Dataset

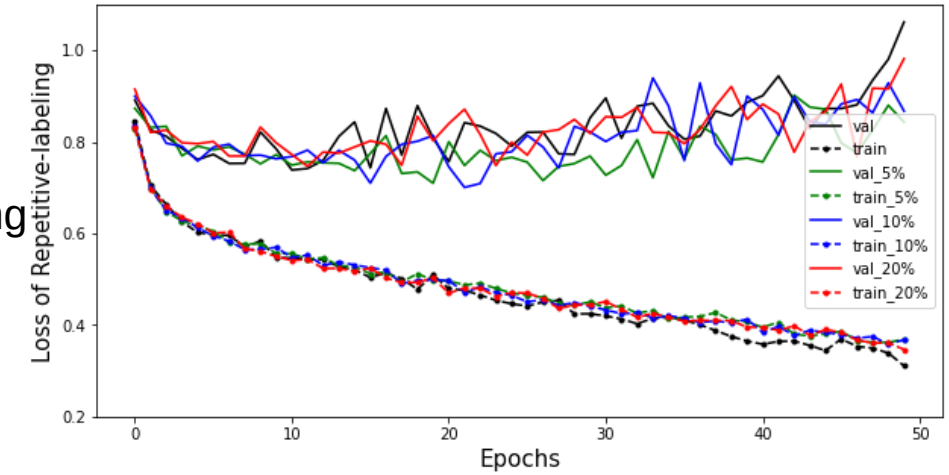
Use EfficientNet CNN model on ISIC Melanoma Detection Dataset with 5%, 10% and 20% redundancy of both

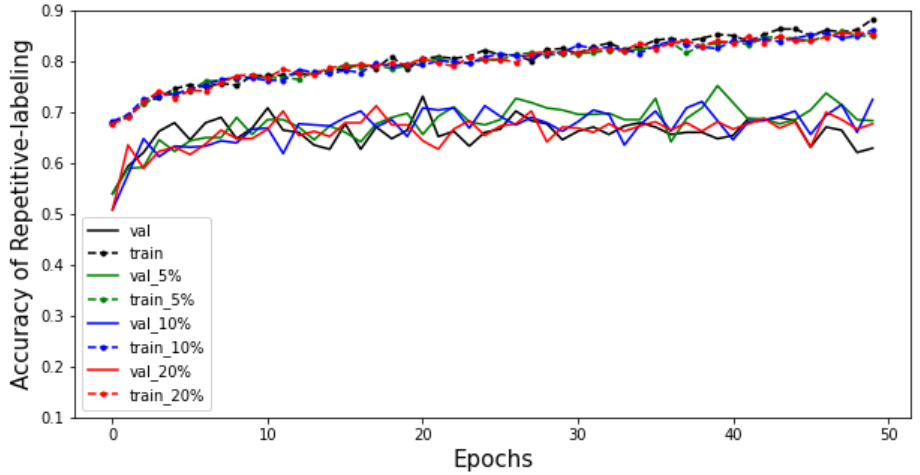
Cross-labelling




Repetitive-labelling




Cross-labelling would potentially damage AI performance, while unclear in repetitive-labelling.

# Future Prospects

- Boost accuracy via pre-processing (e.g., ESRGAN) and post-processing (e.g., Xgboost)
- Manage class imbalance and skin-color bias of AI database and their connection to AI performance
- Incorporate CNN model with metadata ML models

# Selected References:


- “New Zealand skin cancer statistics,” Science Learning Hub, https://www.sciencelearn.org.nz/resources/1329-new-zealand-skin-cancer-statistics (accessed Nov. 18, 2022)
- Narayanamurthy, Vigneswaran et.al., "Skin cancer detection using non-invasive techniques, "*RSC Adv.,* Vol.8, issue 49, pp. 28095-28130, 2018. doi : 10.1039/C8RA04164D.
- “PDQ Adult Treatment Editorial Board, Melanoma Treatment (PDQ®): Health Professional Version, “ PDQ Cancer Information Summaries. https://www.ncbi.nlm.nih.gov/books/NBK66034.1/ (accessed Nov. 18, 2022)
- S. Niyas, S.J. Pawan, M. Anand Kumar, Jeny Rajan, “Medical image segmentation with 3D convolutional neural networks: A survey,” *Neurocomputing*, Vol. 493, pp. 397-413, 2022. [online] Available: https://doi.org/10.1016/j.neucom.2022.04.065.
- Esteva, A., et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, pp. 115–118, 2017, https://doi.org/10.1038/nature21056


Thank you!

Questions?