# Beyond the Surface: Advancing Skin Cancer Detection Across All Tones

*STA 221 Fall 2024*

**Presented By,**

ChiChun Chen
Savali Sandip Deshmukh
Shivani Suryawanshi
Sriharshini D
Yehong Qiu

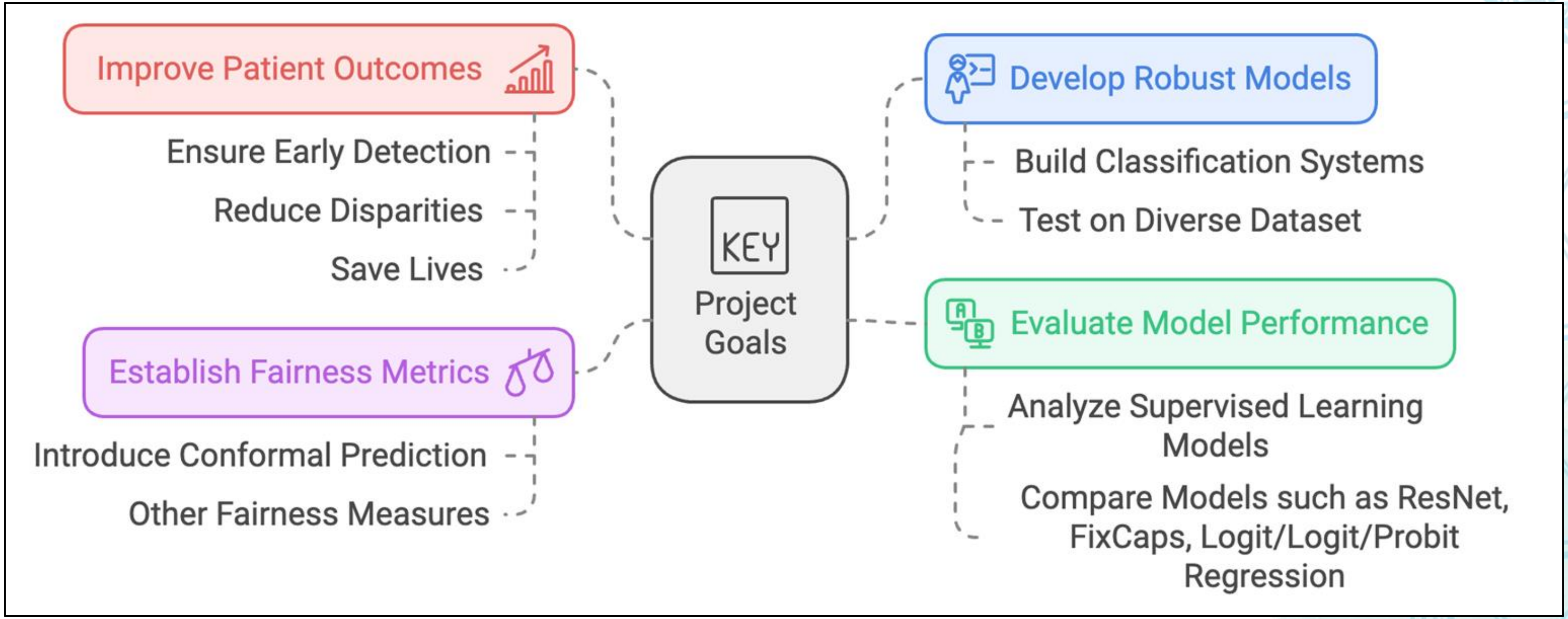
# Introduction

- **Revolutionizing Healthcare**
  - The intersection of AI and medicine is reshaping how we approach diagnostics and patient care.
- **Power of Image Classification**
  - Automated analysis of medical images accelerates diagnostic processes, improving efficiency and precision.
- **Impact on Skin Cancer Detection**
  - Early detection is critical for better outcomes, and AI-driven tools can make this process faster and more accurate.
- **Enhancing Decision-Making**
  - By integrating advanced technologies, healthcare providers can detect abnormalities and make informed decisions more effectively.

# Motivation

- **Highlighting Healthcare Disparities**
  - Current skin cancer detection models are biased toward lighter skin tones, leading to unequal diagnostic outcomes.
- **Raising Awareness**
  - Shine a spotlight on the ethical and scientific importance of inclusivity in medical AI tools.
- **Leveraging AI for Equity**
  - Show how technology can bridge gaps in healthcare access and quality across diverse populations.
- **Call to Action**
  - Inspire innovation in creating AI systems that prioritize fairness, inclusivity, and patient outcomes.

# Goals of the Project

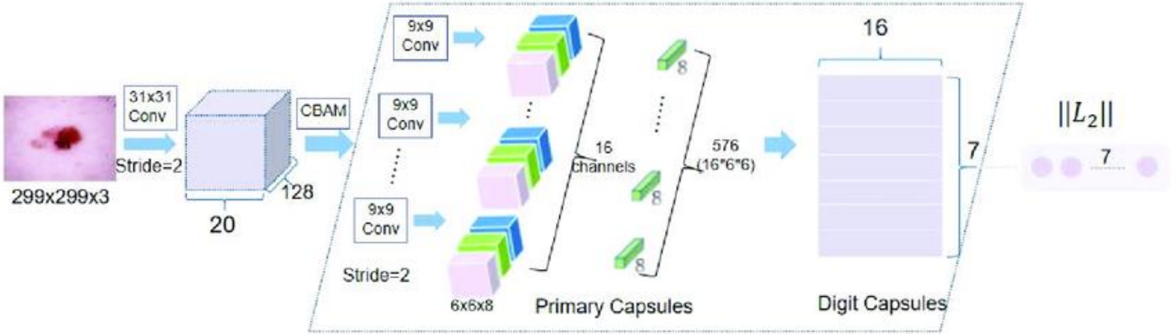


# State of the Arts
# - FixCaps: An Improved Capsules Network

- Advanced version of Capsule Networks (CapsNets) to capture spatial hierarchies in medical images for better generalization in medical imaging.

- Achieved improved accuracy i.e. 96.49% on HAM 10K dataset for skin lesion classification compared to conventional CNNs.

- Mitigates … mented data.



# Literary Review - Fair Conformal Predictors for Applications in Medical Imaging

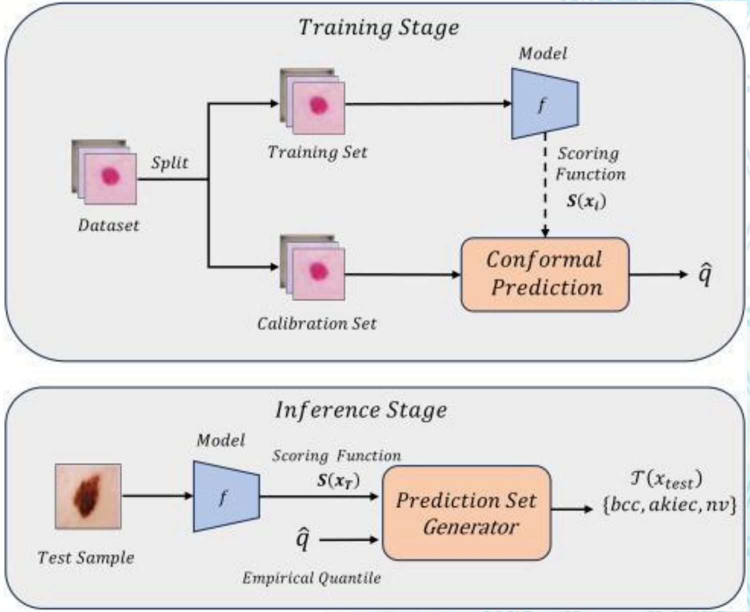
- To **evaluate** uncertainty quantification in medical imaging tasks.

- True label will be included in the **prediction set** with a probability of at least 1−α, achieving **marginal coverage**

$$P\Big(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})\Big) \geq 1 - \alpha.$$

- **Equal confidence levels** across dataset given calibration set is representative of the entire population.

- Can be extended to **Group-Balanced Conformal Prediction**

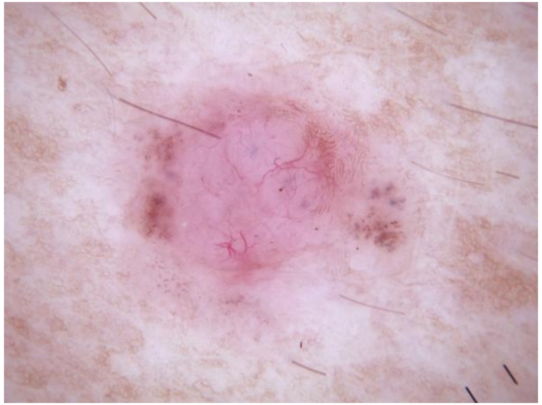


# A demo of Prediction Set

Here is one test sample:

Prediction Set: C={‘bcc’, ‘mel’}

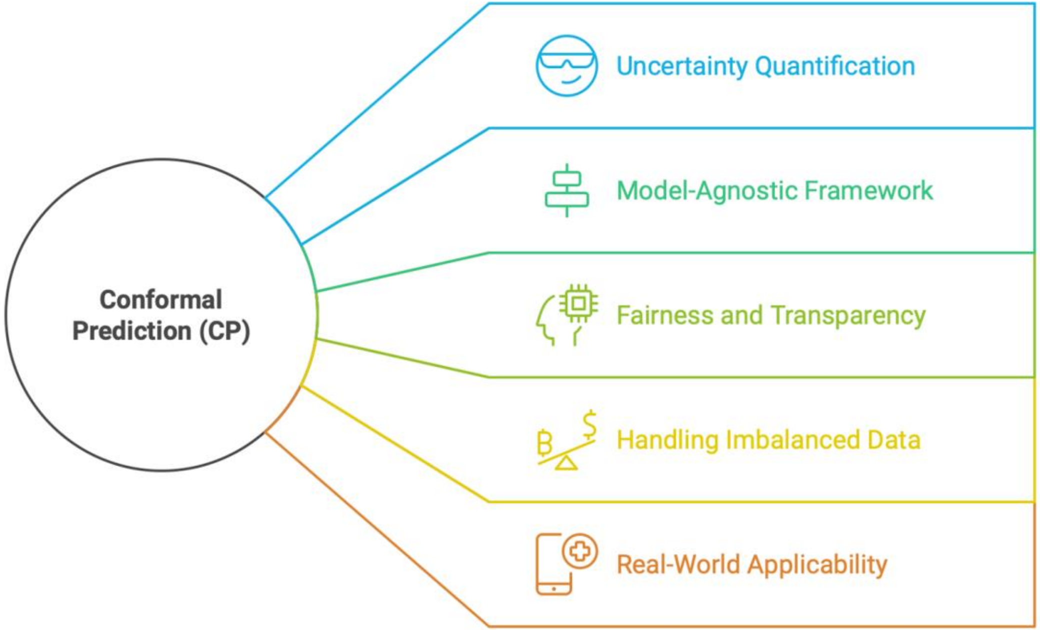
True label: ‘bcc’



True label included!

UCDAVIS

# Why Conformal Prediction

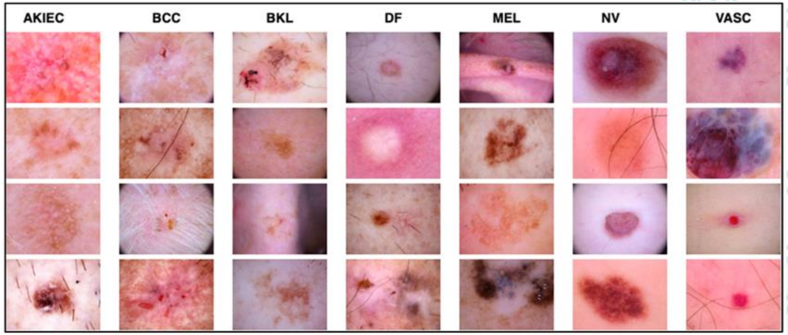



The **trick** is to —

construct a statistically valid **prediction set** based on the **softmax scores** of underlying models

# Dataset

**HAM10000 (**Human Against Machine with 10000 training images**)**: Widely recognized dataset for medical image analysis.
https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
**Size**: 10,015 dermatoscopic images of skin lesions.
**Key Features**:



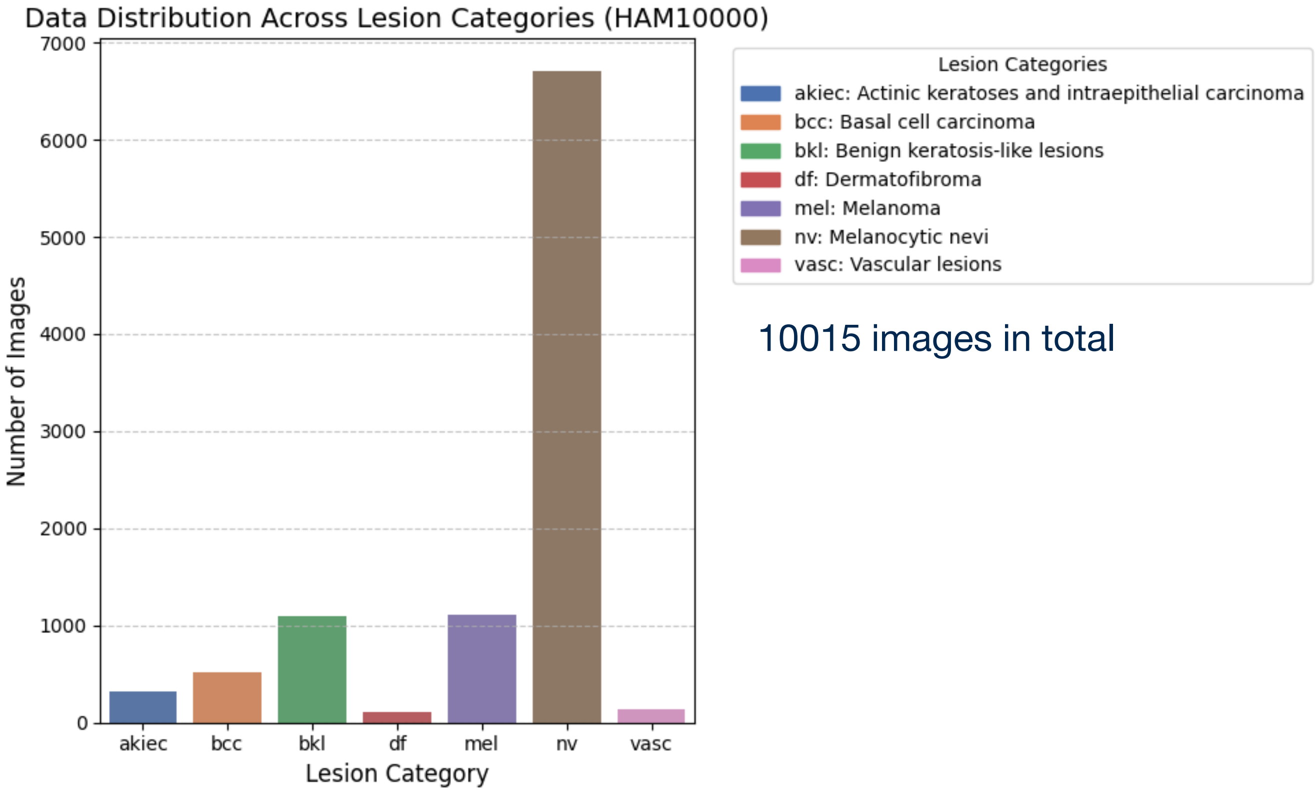

- `imageid`: Unique image identifier.
- `dx`: Diagnostic label (e.g., melanoma, nevus).
- `dxtype`: Diagnostic procedure type (e.g., histopathology).
- `age`, `sex`, `localization`: Patient and lesion details.
- `path`: Image file location.

**Lesion Categories**:

- Melanocytic nevi, melanoma, keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, dermatofibroma.

**Diversity**: Includes images from various skin tones for fair analysis.

# Dataset Distribution




10015 images in total

# Dataset

**Preprocessing:**

- Resize, normalize, and augment images
- Encoded labels to numerical values
- Clean metadata for consistency

**Test set:** ISIC (International Skin Imaging Collaboration) dataset.

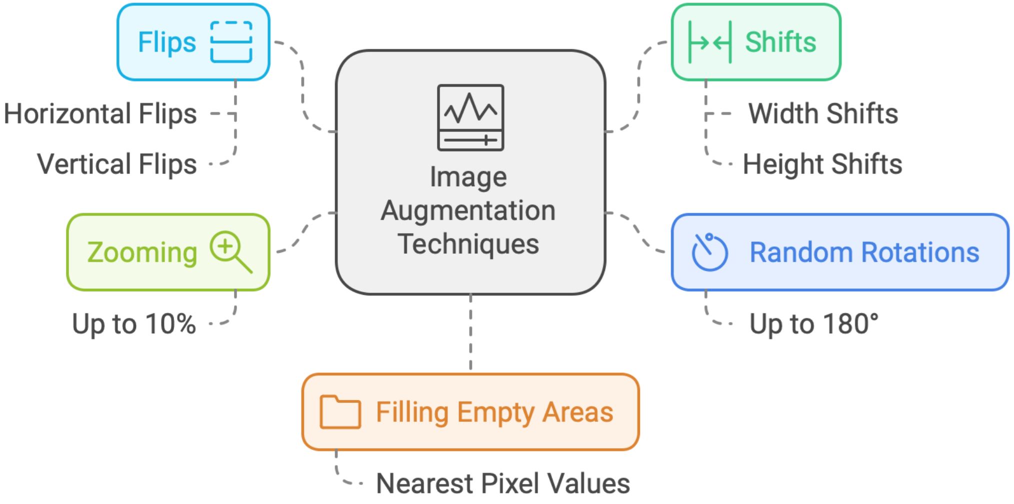
- Serves as a benchmark dataset for skin cancer classification challenges (~240 images)

**Utilization**:

- Images for training,  lesion categories as target labels.

**Significance**: Diverse, well-categorized dataset ensures fairness and robustness in skin lesion classification.

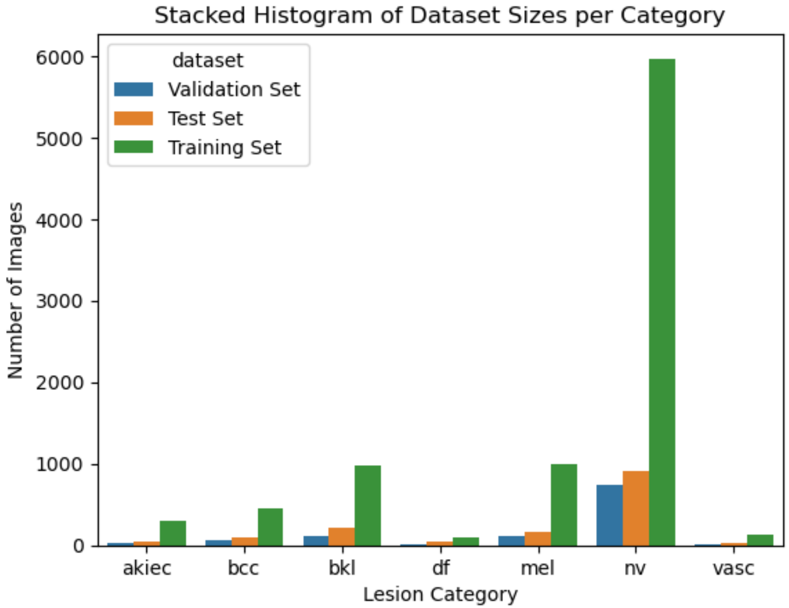
# Dataset Augmentation

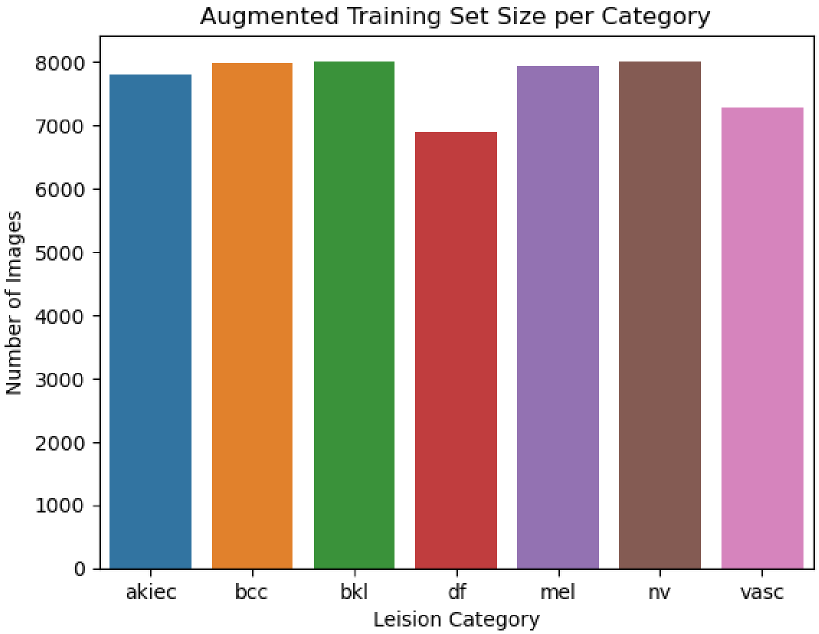



**Final Dataset**:

- 7 lesion categories, each with roughly 8000 images.
- Total dataset size: ~56,000 images.

# Dataset Distribution




UCDAVIS

# Methodology

**Exploratory Data Analysis**
Analyzing the dataset to understand its structure and features

**Data Augmentation**
Applying techniques to balance and enhance the dataset

**Dataset Splitting**
Dividing the dataset into training, validation, and testing sets

**Model Training**
Training machine learning models on the training dataset

**Conformal Prediction Evaluation**
Evaluating model reliability using conformal prediction

**Model Performance Visualization**
Using heatmaps to visualize and interpret model performance

# Logit Regression

**Logit Model:** A regression model used for predicting the probability of a binary or categorical outcome. Uses the logistic function to model the relationship between the dependent variable and independent variables.

**Why Logit for This Analysis?**

**Multinomial Logit:**

- Extended to handle multi-class classification problems.
- Models probabilities for multiple classes using a softmax transformation.
- Provides probabilistic predictions for interpretability.

**Challenges Faced:**

- Computational complexity in high-dimensional data.
- High Dimensional Problem (p>>n): Panelization requires careful hyperparameter tuning via Cross Validation.

# Logit Regression (on PC)

**Preprocessing:**

- Data standardized and encoded numerically for compatibility with the model.
- Principal Component Analysis (PCA) reduced dimensionality to 200 components.

**Data Overview:**

- Dataset includes resized 32×32 normalized images for feature extraction.

**Optimization Techniques:**

- **Hyperparameters:**
  - Regularization applied to the model to control overfitting.
  - PCA components optimized for balanced variance retention.
- **Solver:**
  - Logistic regression trained using the 'lbfgs' solver for faster convergence.

# Logit Regression
# (from PC to GPU Cluster)

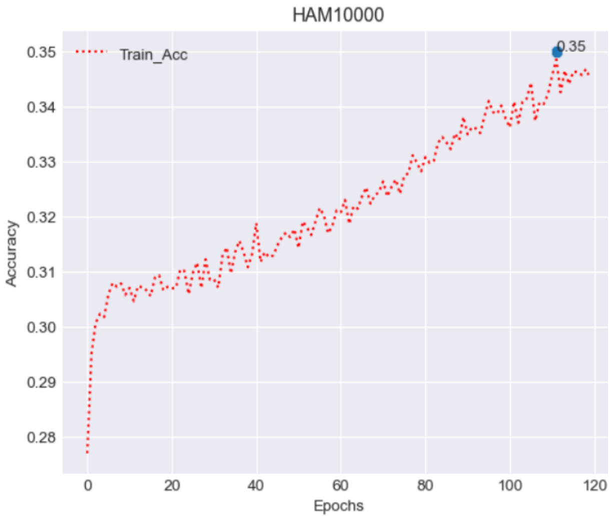
**Drive:**

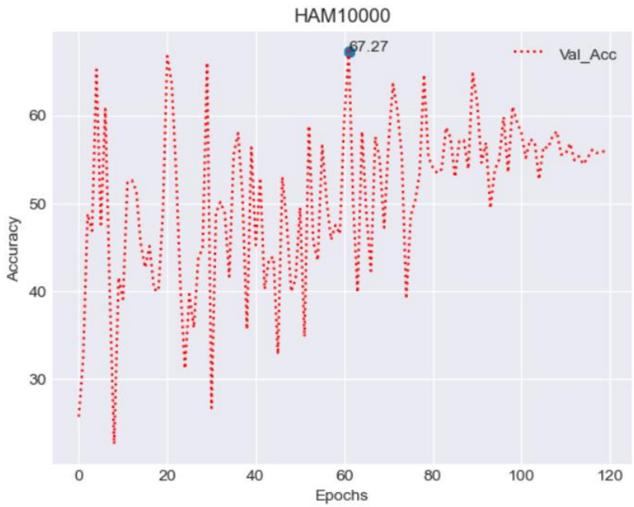
- **Calculation Efficiency:**
  - Roughly 4 hours/epoch on PC.
  - Roughly 1min/epoch on GPU Cluster (1 10GB slice from NVIDIA Ampere A100-80gb GPUs).
- **Multinomial Logit: Single Softmax Layer After Image Flattening**
  - More advanced Deep Learning Models usually use a softmax layer as its last layer.
  - Logit Regression is a good base model to start with.

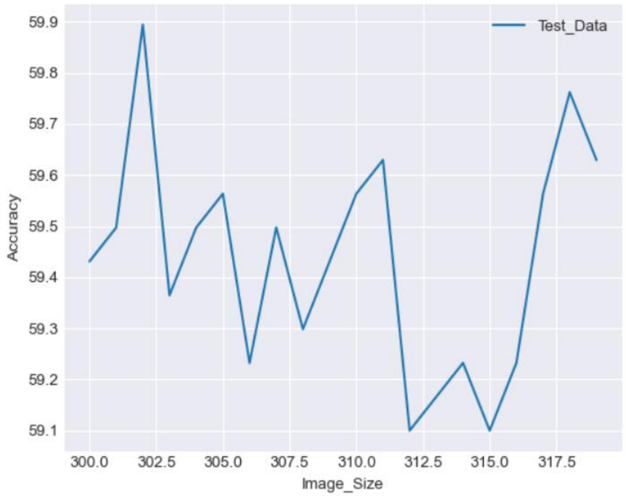
| Model | Params(M) | FLOPs(G) | FPS |
|---|---|---|---|
| LogitRegression | 1.88 | 0.0 | 13797.05 |

# Logit Regression (on GPU Cluster)

Best train accuracy: 35%
Best validation accuracy: 67.27%
Best test accuracy: 59.9%

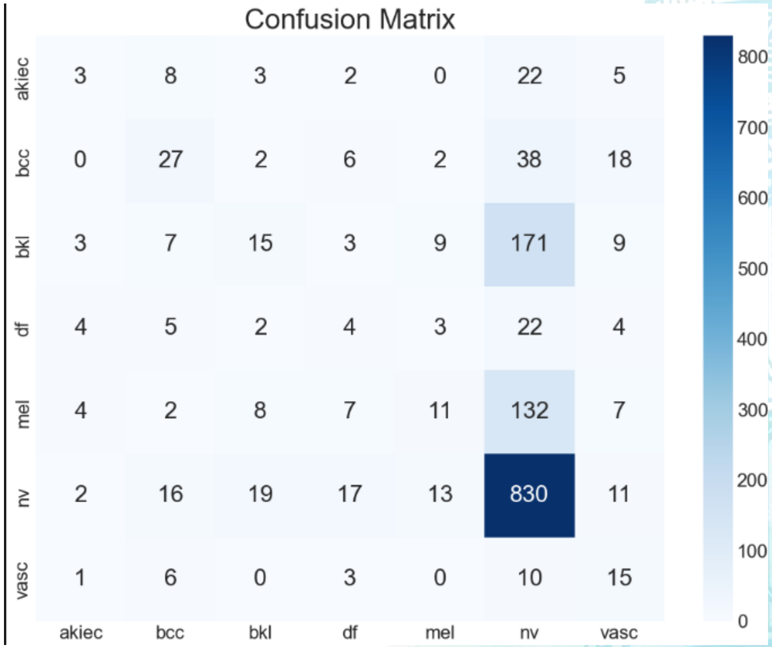


# Logit Regression Test Result (on GPU Cluster)

| Type | Precision | Recall | F1 | Accuracy |
|---|---|---|---|---|
| akiec | 0.1765 | 0.07 | 0.1 | |
| bcc | 0.3803 | 0.29 | 0.329 | |
| bkl | 0.3061 | 0.069 | 0.113 | |
| df | 0.0952 | 0.091 | 0.093 | |
| mel | 0.2895 | 0.064 | 0.105 | |
| nv | 0.6776 | 0.914 | 0.778 | |
| vasc | 0.2174 | 0.429 | 0.288 | |
| Total: | | | | 0.5989 |



# Inception V3 model

**Inception V3 Model:** A deep convolutional neural network (CNN) architecture designed for image classification and feature extraction tasks.

**Why Inception V3 for This Analysis?**

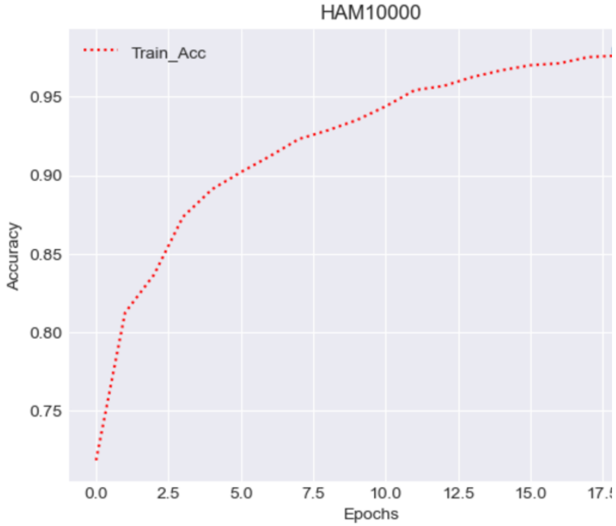
**Key Features:**

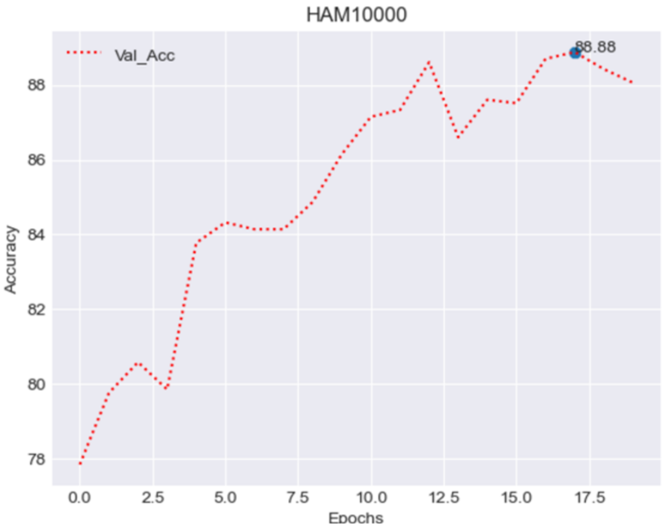
- A Deep Learning model with relatively few parameters (compared to VGG) and DL training techniques (Batch Normalization, Auxiliary classifier,...)
  - ✓ Easy and Stable to train
- Relies on "Inception Modules": blocks designed to capture features at multiple scales
  - ✓ Capture

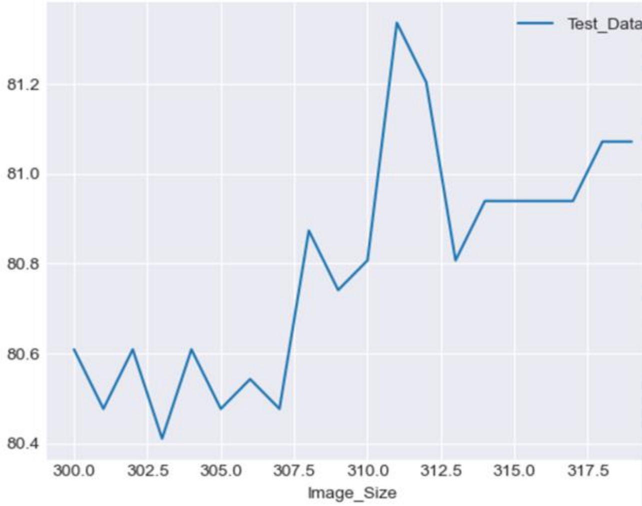
| Model | Params(M) | FLOPs(G) | FPS |
|---|---|---|---|
| InceptionV3 | 21.8 | 5.75 | 142.97 |

# Results

Best train accuracy: 98%
Best validation accuracy: 88.88%
Best test accuracy:81.34%



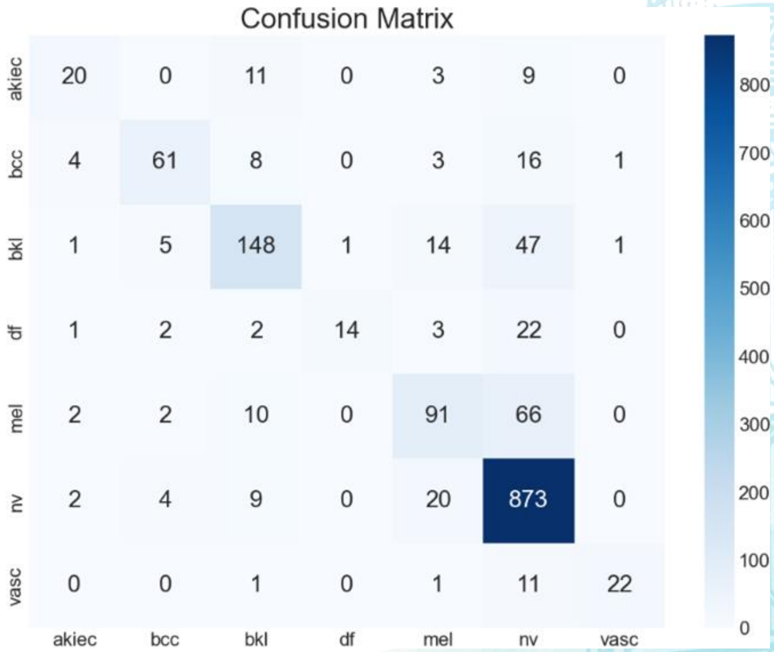




UCDAVIS

# Test Result

| Type | Precision | Recall | F1 | Accuracy |
|---|---|---|---|---|
| akiec | 0.6667 | 0.465 | 0.548 | |
| bcc | 0.8243 | 0.656 | 0.731 | |
| bkl | 0.7831 | 0.682 | 0.729 | |
| df | 0.9333 | 0.318 | 0.475 | |
| mel | 0.6741 | 0.532 | 0.595 | |
| nv | 0.8362 | 0.961 | 0.894 | |
| vasc | 0.9167 | 0.629 | 0.746 | |
| Total: | | | | 0.8134 |



# Model Comparison
# - Conformal Prediction

Marginal Coverage

$$P\Big(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})\Big) \geq 1 - \alpha.$$

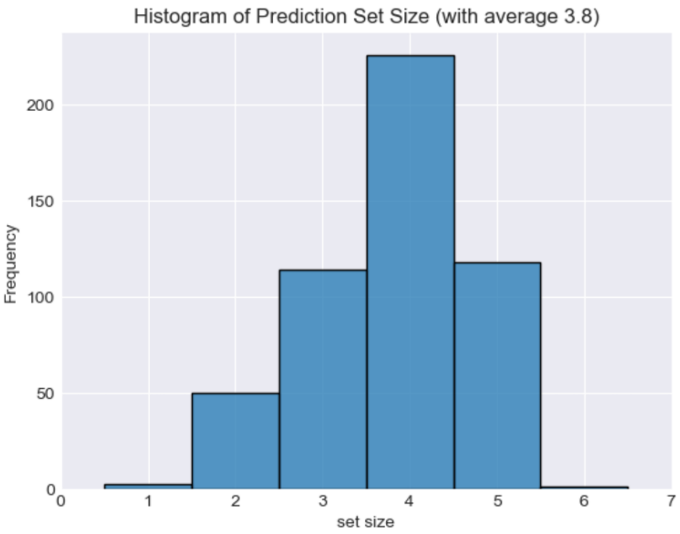
**Rangularized Adaptive Prediction Sets (RAPS) on Validation Set:**

**Experiment Result:**

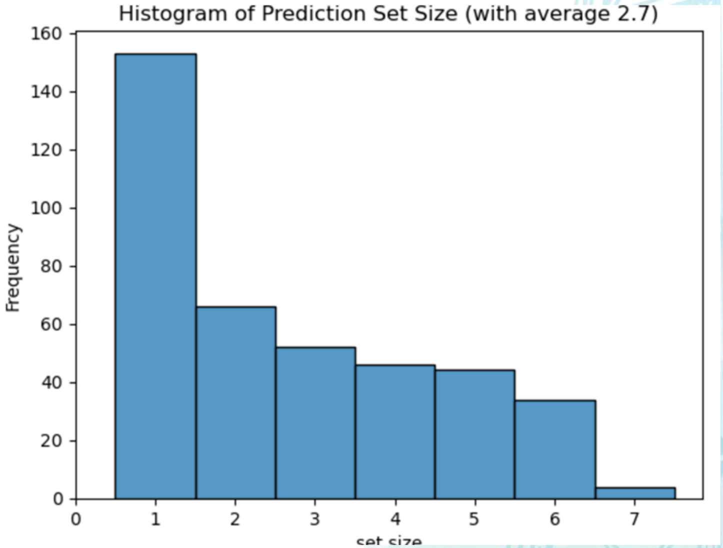
Sig. Level alpha=0.1

n=1000 for Calibration

n=511 for Validation




RAPS Size Distribution for Logit Regression




RAPS Size Distribution for Inception V3

UCDAVIS

# Model Comparison
# - Conformal Prediction

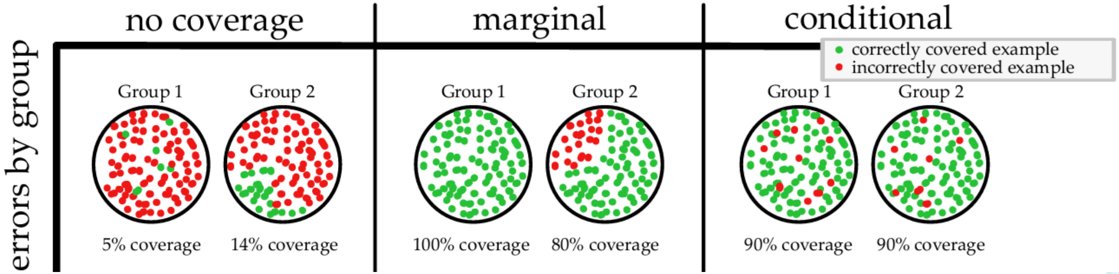
Marginal Coverage

$$P\big(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})\big) \geq 1-\alpha.$$

**Adaptability Evaluation:**

$$\mathbb{P}\left[Y_{\text{test}} \in \mathcal{C}(X_{\text{test}}) \mid X_{\text{test}}\right] \geq 1-\alpha.$$

- **Marginal Coverage** is not adequate for **Conditional Coverage**




- Marginal Coverage is not adequate for Conditional Coverage:
  - ✓ Introduce SSC metric: (S

$$\min_{g\in\{1,\dots,G\}} \frac{1}{|\mathcal{I}_g|} \sum_{i\in\mathcal{I}_g} \mathbb{1}\left\{Y_i^{(\text{val})} \in \mathcal{C}\left(X_i^{(\text{val})}\right)\right\}$$

**Experiment Result:**

Sig. Level alpha=0.1, 1-alpha=0.9

SSC for Logit Regression: 0.5

SSC for Inception V3: 0.83

UCDAVIS

# Model Comparison
# - Conformal Prediction

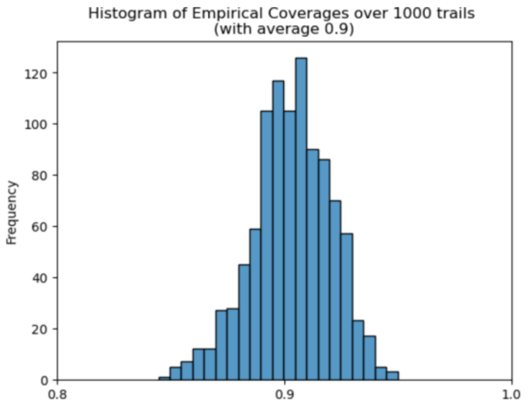
**Correctness Check:** running CP over R trials with new calibration and validation sets, and then calculating the empirical coverage Cj for each: (Average of Cj should center around 1-alpha)

$$C_j = \frac{1}{n_{\text{val}}} \sum_{i=1}^{n_{\text{val}}} \mathbb{1}\left\{Y_{i,j}^{(\text{val})} \in \mathcal{C}_j\left(X_{i,j}^{(\text{val})}\right)\right\}, \text{ for } j = 1, \ldots, R$$

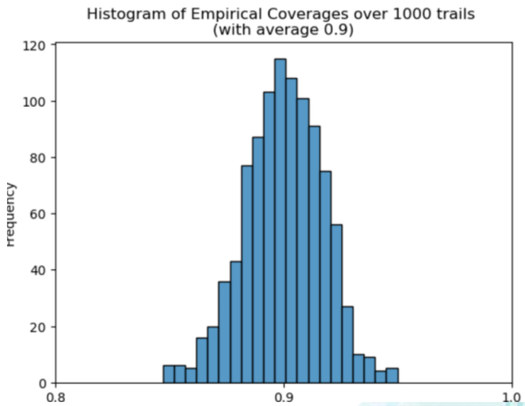
**Experiment Result:**

Sig. Level alpha=0.1

Ran 1000 trails for Logit Regression and Inception V3




Empirical Coverage for Logit Regression




Empirical Coverage for Inception V3

UCDAVIS

# Next Steps and Conclusions

- **Next Steps:**
  - ➢ Resnet : We will train the model from scratch and apply CP to observe results
  - ➢ Fixcaps: the best model known for HAM10K, we will use the network framework and then apply CP
  - ➢ Comparing all these 4 models
- **Conclusions:**
  - ➢ Through this project we successfully implemented and evaluated models using Logit regression and Inception V3, demonstrating initial progress in skin cancer classification.
  - ➢ We hope this project lays a foundation for inclusive, equitable AI in skin cancer detection, bridging gaps in healthcare diagnostics.

# References


[1] M. Li, H. Chen, J. Peng, X. Li, Y. Zhang, and J. Lin, *FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 11, pp. 12165-12173,2022. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9791221

[2] Lu, C., Lemay, A., Chang, K., Hobel, K., Kalpathy-Cramer, J. ,*Fair Conformal Predictors for Applications in Medical Imaging*, Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12008-12016. Available: https://doi.org/10.1609/aaai.v36i11.21459

[3] He, K., Zhang, X., Ren, S., & Sun, J., Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770–778. Available: https://doi.org/10.1109/CVPR.2016.90

[4] Angelopoulos, A. N., Bates, S., A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, arXiv:2107.07511, 2021. Available: https://arxiv.org/abs/2107.07511

# THANK YOU