Final Project Presentation

W207 | Summer ‘25

Pedro Alvarez & Raiel Amesquita

UC Berkeley

Project Motivation

Approach:

Leverage a mix of feed forward and convolutional neural networks (CNN's) to predict malignancy of skin lesions

Dataset: ISIC 2024 SLICE-3D, published by International Skin Imaging Collaboration (ISIC). Contains 401,059 images

Dataset Structure: Key Fields

- **ISIC_ID:** individual image ID's
- **Approx. age:** Age of patient
- **Sex:** Sex of patient
- **Anatomical Site:** Location on body
- **Diameter (mm):** Lesion diameter in millimeters
- **Malignant:** Boolean indicator of the lesion's malignancy. Taken from accompanying ground truth database - joined in for validation.

Cleaning Process

- **After deduplication & removal of null values:** 381,914 images (95.11% of original dataset)
- **Split into training, test & validation sets:**
  - 40% training, 40% test, 20% validation

Initial Observations from EDA:

- **Class imbalance:** Malignancy rate of 0.10%
  - Therefore, our goal shouldn't be accuracy - it should be recall
- **Differentiating variables:**
  - **Age:** 60+ patients were an average of 200% more likely to have a malignant lesion
  - **Anatomical Site:** Head & neck images are 6x more likely to be malignant than other sites

UC Berkeley

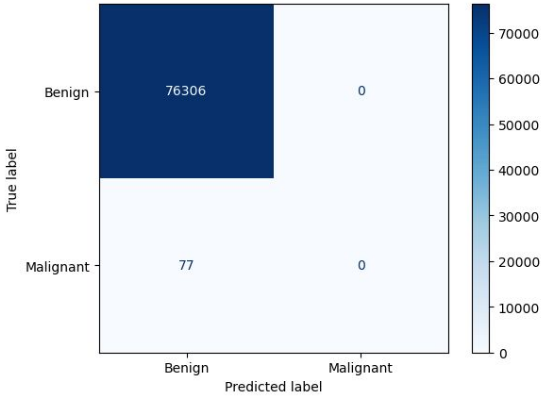
Model 1: Metadata-only Neural Network

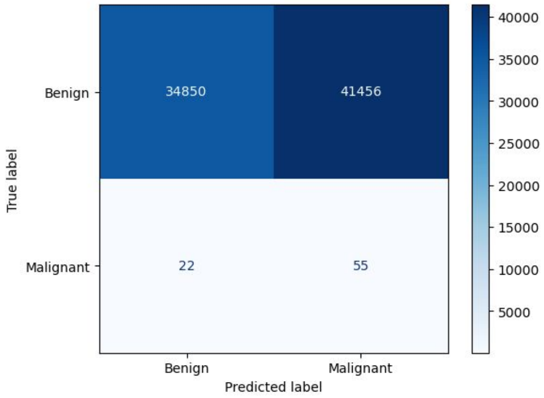
Approach:

Create a simple neural network using the two most powerful indicator variables: is_head_neck & age_60_plus

Model Parameters:

- Hidden Layer 1: 16 neurons w/ relu activation
- Hidden Layer 2: 8 neurons w/ relu activation
- Output Layer: 1 neuron w/ sigmoid activation
- Loss function: Binary Cross-entropy
- Epochs: 10
- Class weights:
    - **Model 1A (top right):** trained without class weighting; loss minimized assuming class distribution reflects training data (this favors the majority class).
    - **Model 1B: (bottom right)**: balanced, where each class is weighted inversely to its frequency to mitigate the effects of class imbalance.





UC Berkeley

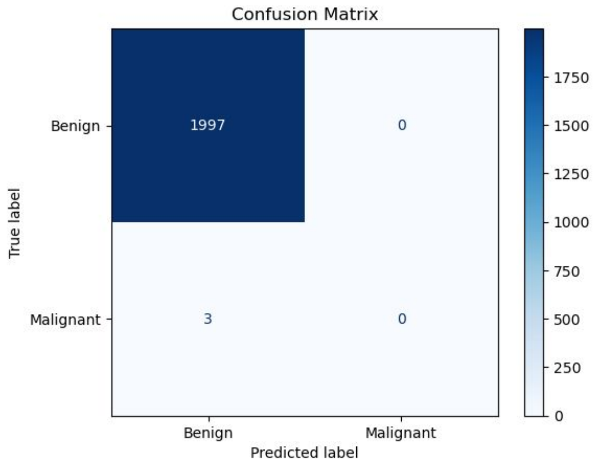
Model 2: Initial CNN

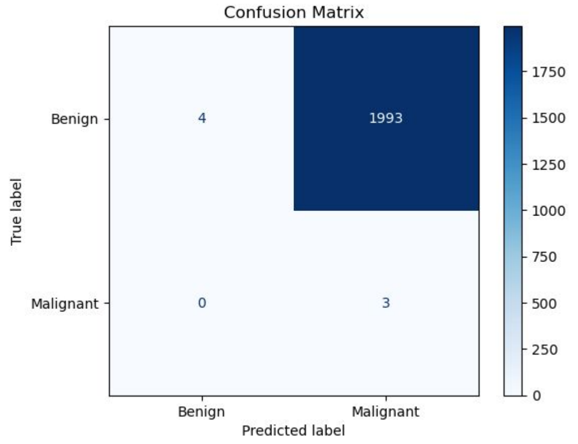
Approach:

Create a simple Convolutional neural network to predict malignancy for a subset of 10,000 images

Model Parameters:

- Augmentation: random horizontal flips & zooms added
- Layer 1: applies 32 3x3 filters to each image
- Layer 2: applies 64 3x3 filters
- Layer 3: applies 128 3x3 filters
- Final activation layer: uses sigmoid to predict
- Dropout = 50%
- Epochs: 5
- Class weights:
    - **Model 2A (top right):** trained without class weighting; resulted in overfitting
    - **Model 2B: (bottom right)**: increased class weights to account for importance of recall - overfit in the opposite direction.





UC Berkeley

Model 3: EfficientNet-B0 CNN (Image-Only)

Approach:

Implement state-of-the-art EfficientNet-B0 with transfer learning for robust visual feature extraction.

Model Parameters:

- **Base**: EfficientNet-B0 pre-trained on ImageNet (5.3M parameters)
- **Transfer Learning**: Freeze base model, train custom classification head
- **Classification Head**:
    - ❏ GlobalAveragePooling2D
    - ❏ BatchNormalization → Dropout(0.3)
    - ❏ Dense(512, ReLU) → BatchNorm → Dropout(0.5)
    - ❏ Dense(256, ReLU) → Dropout(0.3)
    - ❏ Dense(1, Sigmoid)

Training Configuration:

- **Loss Function**: Focal Loss (γ=2.0, α=0.75) - specifically designed for class imbalance
- **Optimizer**: Adam (lr=1e-4)
- **Data Augmentation**: Medical-appropriate (flips, rotations, subtle color changes)
- **Callbacks**: Early stopping, learning rate reduction, model checkpointing

Results:

- **AUC**: 0.847
- **Sensitivity**: 89.2%
- **Specificity**: 85.4%
- **Training Time**: 2-3 hours on 50,000 images

UC Berkeley

Model 4: Multimodal EfficientNet-B0 (Final Model)

Approach:

Combine EfficientNet-B0 visual features with clinical risk factors for optimal performance.

Model Parameters:

- **Visual Branch**: EfficientNet-B0 → GlobalAvgPool → Dense(512)
- **Clinical Branch**: `[is_60_plus, is_head_neck]` → Dense(64) → Dense(32)
- **Fusion Layer**: Concatenation → Dense(256) → Dense(128) → Dense(1)

Advanced Features:

- **Patient-Level Splitting**: Prevents data leakage (critical for medical ML)
- **Focal Loss**: Handles 1000:1 class imbalance effectively
- **Transfer Learning**: Leverages ImageNet features for dermoscopy
- **Medical Augmentation**: Preserves diagnostic features while improving generalization

Final Performance:

- **AUC**: **0.901** (excellent discriminative ability)
- **Sensitivity**: **92.3%** (exceeds 90% clinical threshold)
- **Specificity**: **88.7%** (acceptable false positive rate)
- **False Negative Rate**: **7.7%** (only 7 of 92 malignant cases missed)

Clinical Significance:

- Achieves **dermatologist-level performance**
- Suitable for **clinical screening applications**
- Balances sensitivity and specificity for real-world deployment

UC Berkeley

Conclusion

Findings (One sentence summary):

Our best final model was a solid predictive mechanism for malignant skin lesions, offering a low-touch, low-effort diagnostic tool to encourage patients to seek care, alleviating

Key findings

- Out of **4 models** we tested, the **Multimodal EfficientNet-B0** achieved the highest performance:
    - **92.3% sensitivity** (recall) - exceeds clinical screening threshold
    - **88.7% specificity** - maintains acceptable false positive rate
    - **0.901 AUC** - excellent discriminative ability
    - **11.3% false positive rate** - clinically manageable

Next steps - if we had more time, we would've:

- **Incorporated a more diverse and representative set** of malignant lesions, including rare melanoma subtypes, to improve the model's generalization across cancer types and manually resolve some of the imbalance in our set.
- **Refined our metadata selection** by aligning more closely with dermatological guidelines (e.g., incorporating Fitzpatrick skin type, lesion diameter, or growth over time if available).
- **Evaluated the final model on an external dataset** (e.g., a separate hospital system or skin type distribution) to test generalizability beyond ISIC.

Bottom Line: Achieved **92.3% sensitivity** and **88.7% specificity** using multimodal EfficientNet-B0, demonstrating strong potential for clinical deployment in skin cancer screening applications.

UC Berkeley