IS630
Applied Machine Learning

Group Project:
Skin cancer detection

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Members:
Tran Binh Minh, 01323752
Shwe Tin Aung, 01439396
Aditya Vijay, 01522128
Liu Chih Yuan, 01500507
Lope Immanuell IV Cudiamat Santos, 01520025

1. Background

Problem: Skin cancer, particularly melanoma, is one of the most aggressive and deadly forms of cancer. Accurate, early detection is crucial but currently depends on dermatologist expertise and access to diagnostic equipment.
Compounding Factor:
Rising case numbers + shortage of dermatologists

Proposed Solution: To develop a machine learning model *for image-based skin lesion classification* to distinguish between *benign and malignant (melanoma)* using dermoscopic images. We aim to leverage transfer learning using pretrained CNNs and apply augmentation to improve model generalization and results.

Method	Advantages	Drawbacks
Visual Inspection (Naked Eye)	Simple, widely used	Highly subjective, prone to human error
Biopsy (Histopathology)	Gold standard for diagnosis	Invasive, time-consuming
Reflectance Confocal Microscopy (RCM)	High-resolution, non-invasive	Expensive and limited availability
Optical Coherence Tomography (OCT)	Provides tissue depth information	Limited resolution for pigmented lesions

Usefulness of ML to the Business Problem

Enhanced Feature Recognition
(e.g. identify intricate patterns in skin that may be missed by the human eye

Scalable Diagnostics Deployment
(e.g. enable large-scale skin cancer screening via various platforms

Accessible and Non-Invasive Detection
(e.g. Model require only a photo of the lesion, making early detection more accessible in rural areas

Continuous Improvement
Model accuracy improves with more labeled data over time, and may be applicable to other diseases

2. Literature Review

No	Study name	Key Methodology/Model(s)	Feature engineering approaches	Improvement gaps
1	Venugopal et al. (2023)	DNN (EfficientNetV2-M), Fine-tuning	- **Transfer learning** leverages features from pre-trained model. - **Data augmentation techniques** (Geometric transformations: Imagine Rotation..., color variations) used to expand dataset to solve class imbalance .	- Data scarcity of malignent cases leads to susceptible to data outliers. - **Data quality issues** (such as duplication or multiple shots of the same case)
2	Rahman et al. (2021)	Ensemble (DenseNet, ResNet, Xception, SeResNeXt, ResNeXt)	- Deep learning models automatically learn features from raw data - **Ensemble learning combines features/predictions from multiple models**	- This ensemble approach increases the computational complexity and resource demand due to the integration of multiple deep learning models
3	Gouda et al. (2022)	Custom CNNs, Pre-trained (ResNet50, InceptionV3, Inception ResNet)	-**ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) is used for data augmentation** to generate synthetic data. - Deep learning models automatically learn and extract relevant visual features.	- Generalizability to diverse patient populations (skin tone) and varying imaging conditions (imagine resolution, angles) remains a concern. - Difficulty in understanding how deep learning models reach conclusions ("black box" nature). Potential biases present in the training data.
4	Su et al. (2024)	Self-Transfer GAN (STGAN), T-ResNet50 with Test-Time Augmentation (TTA)	STGAN generates 256×256 images by first training on all classes for global knowledge, then fine-tuning per class. Integrates Barlow Twins and Freeze-D in training; TTA applied during inference.	Generated images inherit natural artifacts (e.g., hair, veins) that may hinder classification, authors **propose using image inpainting in future to remove artifacts before training**
5	Oumoulylte et al. (2023)	CNNs, Transfer Learning (VGG16, VGG19)	Transfer learning is used for feature extraction. Deep learning models automatically learn features.	**Deep neuron network models makes it difficult to understand the rationale behind the model's predictions,** highlighting the need for Explainable AI (XAI).
6	Nawaz et al. (2025)	FCDS-CNN (custom CNN), ResNet152V2, EfficientNetV2B0, InceptionResNetV2, MobileNetV3	Resizing, normalization, class weighting (inverse frequency), data augmentation (rotation, flip, zoom), transfer learning, **batch normalization, dropout layers**	Pre-trained models had difficulty with minority class differentiation; **high computational cost in deep models;** potential for further fine-tuning and integration of interpretability tools (e.g., Grad-CAM)
7	Ha et al. (2020)	Ensemble (EfficientNet B3-B7, SE-ResNeXt101, ResNesSt101	Data augmentation using Albumentations: flip, rotate, blur, noise, distortion, cutout, hue/saturation shift, etc. Metadata goes through fully connected layers before concatenation.	Imbalanced dataset (1.76% positive), label mismatch across years, noisy public LB evaluation. Metadata integration improves diversity but may reduce standalone performance. **Challenges in generalizing models across different data distributions.**

3. Literature Review Recap

Takeaways for model development

- **Leverage pre-trained model** (EfficientNet) and fine-tune to suitable with our problem.
- **Use ensemble learning to combine predictions from multiple models;** explore strategies to distribute weights among them (no definitive best ratio).
- **Consider GAN-based techniques (e.g. for early melanoma detection or class imbalance handling).**
- **Apply batch normalization and dropout layers for better model regularization.**
- Use **data augmentation** (rotation, flipping, zooming, translation, color variation) to increase diversity and address class imbalance.
- Apply **ESRGAN** (Enhanced Super-Resolution GAN) to boost image resolution, especially useful post-augmentation.
- Apply **image inpainting** to remove noise or irrelevant artifacts:
  - **Exemplar-Based Inpainting** (Criminisi et al., 2004)
  - **PatchMatch** (Barnes et al., 2009) for efficient patch matching.

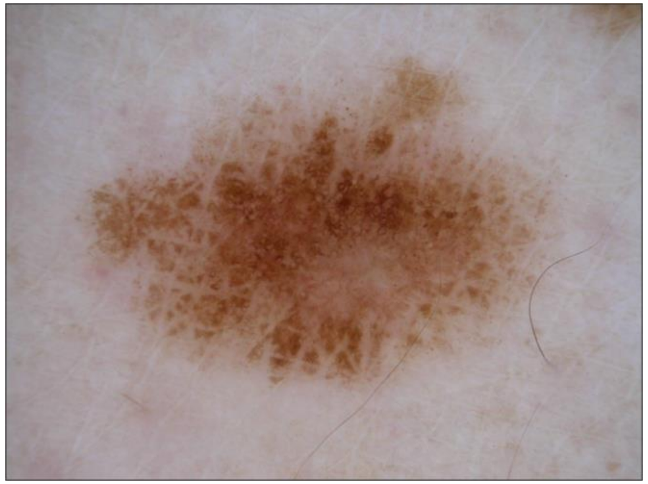
Previous Research Problems & Our Proposed Solution To Address

- **Class Imbalance (malignant / benign)**
  → Data augmentation to boost malignant samples.
- **Cross-Population Generalization**
  Poor performance on darker skin tones and acral lesions.
  → Skin-tone transformation via VGG-19 style transfer to create darker-tone images while preserving pathology.
- **Noise: low resolution, hair, varied angles**
  → Filter out noise objects with image-inpainting algorithms.
  → Flip / rotate images to a consistent orientation before training.
- **DL model opacity**
  → XAI methods (Saliency Maps such as Grad-CAM, SmoothGrad) to reveal key image regions influencing predictions.

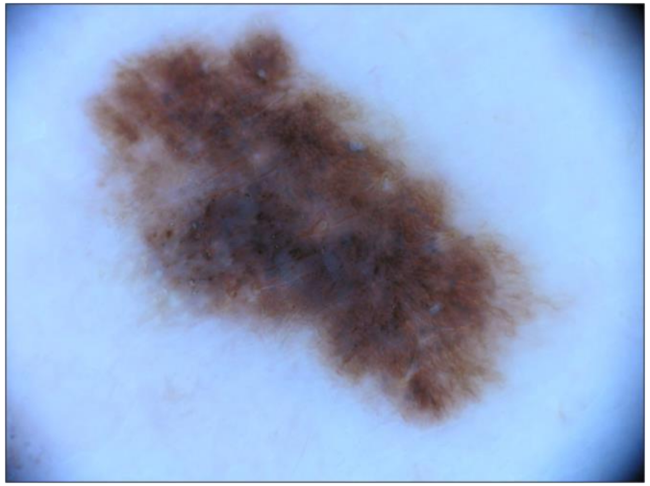
4. Dataset Description

The ISIC (International Skin Imaging Collaboration) Dataset is a large, publicly available collection of over **33,000 dermoscopic skin images** with rich metadata including diagnosis, patient age, sex, and lesion location. This dataset supports **multiple classification tasks** such as distinguishing between benign and malignant lesions, identifying specific skin cancer types, and segmenting lesion regions. It is widely used for **skin cancer prediction and diagnosis research**

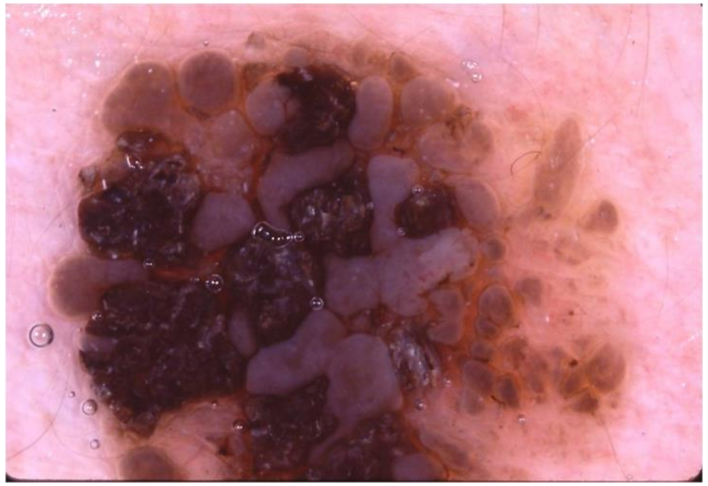
Example of the dataset



Benign tumor

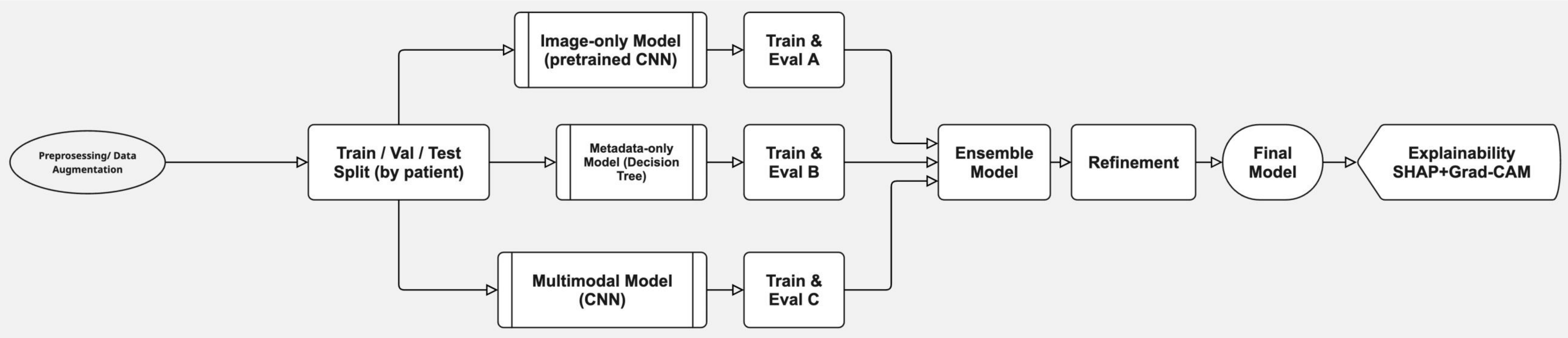


Benign tumor



Malignant melanoma tumor

5. Execute Plan



Phase 1: Data Preparation (Pre-processing + Train test split)
•**Prep inputs:** Image preprocessing including resizing, normalization and format conversion.
- **Safe split:** divide cases (not files) into train/val/test—reuse splits for all experiments.

Phase 2: Model Training, Evaluation and Ensemble
- **Train 3 Baseline models:** fine-tuned CNN models on images, decision-tree & ensemble models on metadata, and a fused CNN.
- **Ensemble selection:** Combine predictions from all three models using weighted voting/ stacking.

Phase 3: Refinement and Optimization
- **Hyperparameter tuning**: Learning rate scheduling and regularization tuning.
- **Architecture tweaks**: Unfreeze more layers of the pretrained CNN or add batch-norm/ extra fusion layers for stability

Phase 4: Final model and Explainability
- **Images:** Potentially explore applying Deep SHAP (via SHAP's DeepExplainer) and Grad-CAM to highlight pixels driving each CNN decision
- **Deliverable:** Binary prediction and potential model explainability.

6. Evaluation Metrics Selection: Clinical Necessity & Literature Alignment

Why These Three Metrics Are Critical for Skin Cancer Detection?

1. Recall (Sensitivity) - Patient Safety Priority

- **Clinical Imperative:** Missing a malignant case can be fatal - false negatives are unacceptable in cancer detection
- **Literature Evidence:** Su et al. (2024) achieved 88.85% sensitivity, demonstrating this as a key benchmark
- **Target:** >90% recall to ensure comprehensive cancer detection
- **Medical Relevance:** Dermatologists prioritize catching all suspicious cases over reducing false positives

2. Precision - Healthcare Efficiency & Patient Well-being

- **Resource Management:** Reduces unnecessary biopsies, procedures, and patient anxiety from false positives
- **Literature Support:** Su et al. (2024) reported 90.23% precision as a balanced approach
- **Target:** >90% precision to maintain clinical credibility and minimize healtahcare burden
- **Patient Impact:** Prevents psychological distress and unnecessary medical interventions

3. AUC-ROC - Overall Discriminative Power

- **Threshold Independence:** Evaluates model performance across all decision thresholds, crucial for clinical deployment
- **Literature Benchmark:** Multiple studies report AUC as primary metric (Ha et al.: 0.96, Nawaz et al.: 0.97)
- **Target:** >0.96 AUC to match state-of-the-art performance
- **Clinical Decision Support:** Enables flexible threshold adjustment based on clinical context (screening vs diagnosis)

Literature-Driven Justification Addressing Key Improvement Gaps from Literature:

Class Imbalance Challenge: Our metrics directly address the 1.76% melanoma prevalence noted in Ha et al.

Subject Manual Assessment: AUC-ROC provides objective performance measurement beyond human subjectivity

Generalizability Concerns: These metrics are consistent across diverse populations and imaging conditions

"Black Box" Nature: Precision and recall provide interpretable performance indicators for clinical acceptance

These three metrics provide a comprehensive, clinically -relevant evaluation framework that addresses both patient safety (recall) and healthcare system efficiency (precision), while ensuring robust discriminative performance (AUC-ROC) across all operational thresholds.

References

1. **Venu Gopal et al. (2023):** "An Integrated Deep Learning Model for Skin Cancer Detection Using Hybrid Feature Fusion Technique." (Based on the arXiv preprint with similar title and year).
2. **Rahman et al. (2021):** "A Method for Detecting Skin Cancer Disease Based on Deep Learning in Dermoscopic Images."
3. **Gouda et al. (2022):** "Skin Cancer Diagnosis Utilizing Hybrid Discrete Cosine Transform and High-Performance Convolutional Neural Networks."
4. **Su et al. (2024):** "Early Stage Skin Cancer Detection Using Deep Learning: A Comprehensive Model for Improved Treatment Outcomes and Survival Rates."
5. **Oumoullyte et al. (2023):** The provided information is limited. It's possible this refers to a study using CNNs and transfer learning for skin feature extraction, but a specific publication title is not evident from the table.
6. **Nawaz et al. (2025):** "Optimizing Deep Learning for Skin Cancer Classification: A Computationally Efficient CNN with Minimal Accuracy Trade-Off." (Based on the arXiv preprint with similar title and year).
7. **Ha et al. (2020):** "Skin Cancer Disease Detection Using Transfer Learning Technique."