BTI Final Presentation
A.Y. 2025-26
Project Title: A Comprehensive Review on Skin Lesion Analysis and Prognostic Modeling using Deep Learning

Presented by : C069 Raj Shukla
C080 Aalia Fatema Dandawala
C096 Vedant Jadhav
Under the guidance of : Dr. Prashasti Kanikar

Roadmap

- Introduction
- Problem definition
- Literature Review / Market Survey
- Proposed System/ Algorithms / Architecture
- Design Diagrams
- Implementation
- References

Introduction

Background:

Skin cancer is one of the most common cancers worldwide.
Melanoma, while less common, is the deadliest form.
Early detection can increase 5-year survival rates from ~25% (late stage) to >95% (early stage).

Current AI Solutions:

- Classify lesions from static dermoscopic images (CNN, EfficientNet).
- Segment lesion boundaries for measurement.
- Generate synthetic images for dataset balancing using GANs.

Gap: These models work on "now" images — they do not predict how lesions will look in the future.

Problem definition

- Despite advances in dermatology-focused AI, existing systems are limited to detection and classification of lesions at a single point in time.
- They lack the capability to forecast how lesions may evolve, especially when longitudinal patient records are unavailable.
- As a result, malignant transformations often remain undetected until they become visually apparent, significantly lowering patient survival rates.
- There is a need for AI-driven prognostic models that can provide visual predictions of lesion progression, enable earlier intervention and improve clinical outcomes.

Literature Review

Citation	Accuracy	Dataset	Methodology	Findings	Limitation
[1]	90–92% (Dice)	ISIC 2017/2018	Mamba-powered UCM-Net with CNN-MLP-Mamba blocks, group loss function	Lightweight MUCM-Net achieves 0.90–0.92 Dice, 0.94 sensitivity, 0.97 specificity; mobile-efficient (0.055–0.064 GFLOPs)	Limited validation on non-Caucasian skin tones; computational constraints for real-time use
[2]	83.8% (Balanced)	ISIC 2019	XAI with ABELE, progressive growing AAE for exemplars/counterexemplars	Boosts clinician confidence by 10–20%; RMSE 0.08–0.24; demonstrates clear class separation in latent space	Lack of standardized XAI metrics; limited to specific lesion classes
[3]	- (RMSE 0.08–0.24)	ISIC 2019	AAE for exemplars/counterexemplars, progressive growing to avoid mode collapse	Enhances interpretability; aids misdiagnosis correction	Limited to classification; requires diverse data validation
[4]	73.7%	Interactive Atlas of Dermoscopy	Multi-task CNN for direct management prediction (EXC, CLNC, NONE) + 7-point criteria	13.73% accuracy gain over inferred diagnosis; 24.56% fewer over-excisions; achieved AUROC 0.844	Limited dataset size; needs broader clinical validation
[5]	92–96%	Public (e.g., ISIC, DermNet)	CNN-SVM ensemble with GAN augmentation, CapsNet for spatial features	Outperforms standalone CNN (88.83%) and CapsNet (90.15%) by $3-5\%$	GAN training instability; limited to dermoscopic images
[6]	94.5% (Dice 90.1%)	ISIC 2018	EGAN/MGAN with squeeze-excitation encoder, morphology-based smoothing loss	EGAN: 90.1% Dice, 83.6% Jaccard (SOTA segmentation); MGAN: 2.2M params, 13 FPS (mobile-efficient)	Synthetic data realism varies; high computational cost for EGAN
[7]	- (pAUC 0.1755)	ISIC 2024 SLICE-3D: 401,059 img	ViT hybrids (EVA02, EdgeNeXtSAC), GBDT ensemble, Stable Diffusion synthetics for augmentation	Achieved highest pAUC (0.1755); notably enhanced robustness against class imbalance/rare malignant cases	Training instability in diffusion models; needs diverse skin tone data
[8]	88.5% (Balanced)	ISIC 2018 (12,500+ images)	Benchmark for segmentation, attribute detection, classification (using Balanced Accuracy)	Top segmentation Jaccard 0.838; established generalized testing on internal/external data	Generalization issues (over 10% failure rate for some models); limited dermoscopy data
[9]	- (5–15% gain)	Various (e.g., ISIC)	Review of GANs for augmentation/segmentation	GANs consistently improve accuracy by $5-15\%$ by addressing data scarcity/imbalance	GAN instability; ethical concerns with synthetic data
[10]	86.57% (Accuracy)	Life 2022 dataset	MobileNetV2 + LSTM for multi-class prognosis and stateful prediction	High efficiency in stateful prognosis; achieved 86.57% accuracy with combined MobileNetV2-LSTM model.	Limited performance metrics; potential data imbalance issues
[11]	0.8601 (AUC)	Discovery Dataset, ISIC 2018	Two-stage pipeline (Faster-RCNN + ResNet-50) for lesion identification and malignancy prediction	Image model performs comparably to expert dermatologists; visual data significantly outweighs clinical data alone.	Limited dark skin tone samples; discovery dataset is proprietary
[12]	89.47% (Acc)	TCGA-SKCM (Genomic Profiles)	SVC-W (Support Vector Classification with Weight), feature selection (SVC-L1, WEKA-FCBF) on multi-omics data (mRNA, miRNA, methylation)	17 mRNA expression features distinguish primary/metastatic tumors with 0.95 AUROC; mRNA most predictive omic layer.	Limited external validation of genomic markers; model relies on RSEM values from RNAseq
[13]	91.17% (Acc)	HAM10000	Ensemble of CNNs (DenseNet169/121, ResNet101) for classification, data balancing, transfer learning	DenseNet121 achieved highest single-model accuracy (91.17%). Hybrid models demonstrated efficacy.	Models were sensitive to data split ratio; generalization outside HAM10000 is not guaranteed
[14]	0.9426 (AUC)	HAM10000, ISIC 2017	Multimodal fusion using ALBEF (Vision Transformer + BERT) to combine images and textual metadata (age, sex, location)	Multimodal model significantly outperforms image-	Only basic textual features (age, sex,

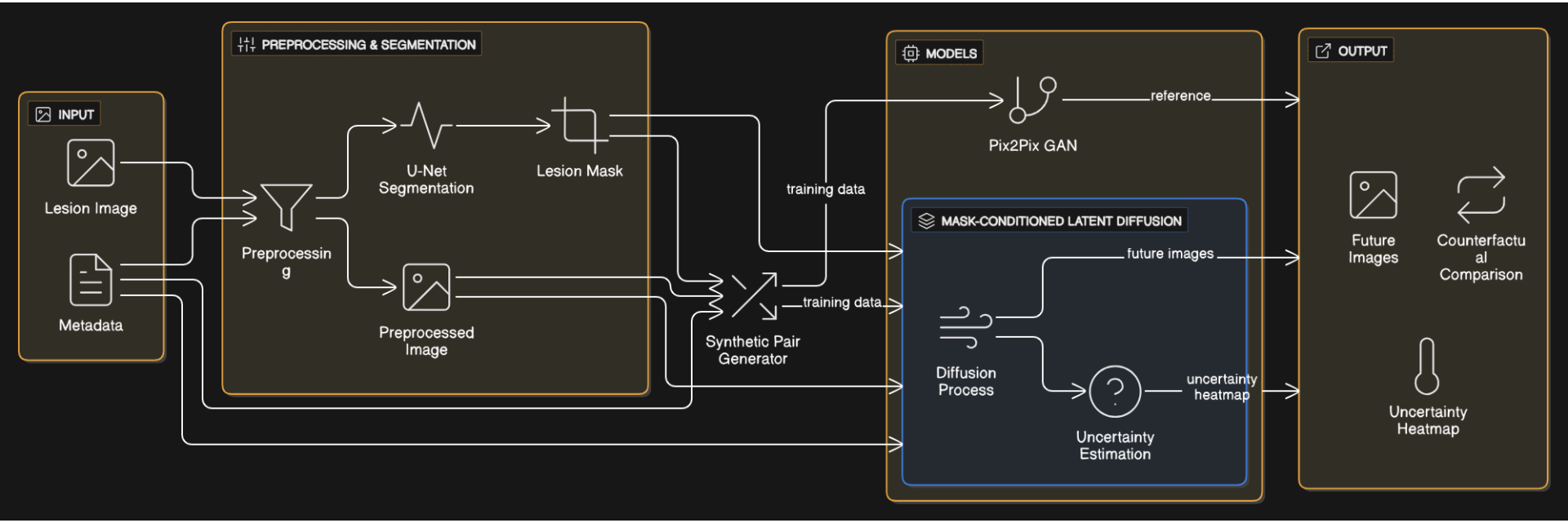
Literature Review

[15]	93% (Acc)	DermNet (Multiclass)	UA-Network + DCNN + LSTM + MRFs (Markov Random Fields) for multi-class lesion classification and prognosis	System evaluates if a lesion is "responding to treatment or becoming severe"; LSTM tracks sequential patient history (prognosis).	Relies heavily on structured patient history being available; validation primarily on DermNet
[16]	93.5% (Acc)	HAM10000	Ensemble of VGG16, ResNet50, ResNet101, explicitly engineered data augmentation (image transformations)	Ensemble significantly outperformed individual models and prior work on the same dataset; increased performance on less frequent classes.	CGAN implementation failed to produce usable synthetic images; reliance on simple image transformations.
[17]	75.6% (Acc)	HAM10000, D7P	LesionGen (Concept-Guided Diffusion Model) fine-tuned with rich, structured dermatological captions	Achieved superior overall accuracy and notable gains in worst-case performance for rare classes (e.g., DF); concept-guided prompts enhance generation quality.	Performance still falls short of models trained entirely on real data; relies on pseudo-generated captions for HAM10000.
[18]	68.11% (Top-1 Acc)	ISIC2019-C, Dermnet-C (Corrupted Benchmarks)	Diffusion Model Driven Test-Time Image Adaptation (TTA) with Structured Guidance and Confidence-based Self-Ensembling	TTA significantly enhances robustness against image corruptions (blur, noise) by adapting corrupted input to the source domain.	Requires a separately trained diffusion model; computational cost may be high for real-time edge deployment.
[19]	0.8966 (DSC)	ISIC 2018, PH2, HAM10000	DermoSegDiff: Diffusion Model with Boundary-Aware Loss Function	Boundary-aware loss prioritizes perimeter pixels, achieving SOTA segmentation performance against	Requires high-quality ground truth masks for boundary loss calculation; high
[20]	0.8122% (Acc)	HAM10000	Custom CNN with Dropout/BatchNormalization; ensemble of deep networks (ResNet50+DenseNet121+InceptionV3)	DenseNet-121 achieved highest single-model accuracy (91.17%). Ensemble approach improved performance but only reached $\sim 81\%$.	Accuracy highly dependent on proper data balancing; complex hybrid model did not guarantee superior performance.
[21]	0.866 (AUC)	Dermoscopy/Clinical (Proprietary)	Hybrid Multimodal CNN (ResNet50/InceptionV3/VGG16) for fusion of macroscopic, dermoscopic, and clinical metadata	Multimodal fusion substantially boosts AUC (up to 0.866) and specificity. Outperformed single-image prediction models.	Dataset size is limited; limited generalization to different clinical settings.
[22]	0.78 (Acc)	Proprietary (Wide-Field/Dermoscopy)	Two-stage framework (Faster-RCNN for detection + ResNet-50 for malignancy prediction)	Framework effectively detects lesions in complex wide-field images; achieved comparable AUC to experienced dermatologists.	Proprietary nature limits reproducibility; small dataset size.
[23]	0.77 (Acc)	Proprietary, ISIC 2018	Ensemble of two customized CNNs (InceptionV3/ResNet-152) trained with auxiliary tasks (e.g., shape features)	Multi-task loss (diagnosis + auxiliary features) significantly improved performance and feature alignment.	Auxiliary feature definition is heuristic; complex training pipeline.
[24]	0.89 (Acc)	HAM10000	ResNet50 with transfer learning, fine-tuning, and explicit data augmentation	Achieved strong accuracy by applying transfer learning to handle diverse lesion categories.	Suffered from dataset imbalance (Nevus majority); a multi-step process was required.

Literature Review

[25]	0.92 (Acc)	HAM10000	Ensemble of multiple pre-trained CNNs (VGG16, ResNet101/50), explicit image transformations for data augmentation	Ensemble method compensated for single-model weaknesses; explicit augmentation was deemed sufficient when GAN synthesis failed.	CGAN synthesis failed to generate realistic images for optimal augmentation; generalization depends solely on image transformations.
[26]	0.9411 (Acc)	HAM10000, ISIC 2017	ALBEF (Multimodal Fusion Model) combining images and text (age, sex, location)	Demonstrated that fusing images and text is superior (up to 0.9411 Acc, 0.9426 AUC) for generalized skin cancer classification.	Only used basic textual metadata; did not validate on diverse skin tones.
[27]	0.8163 (Acc)	Primary Tumors (SKCM-TCGA)	Threshold-based logistic regression classifier using mRNA expression data to distinguish early vs. late-stage primary tumors	Identified 37 key mRNA features; successful early vs. late-stage tumor stratification.	Only useful for classification, not visual prognosis; retrospective data.
[28]	0.8876 (Acc)	Primary/Metastatic (SKCM-TCGA)	Random Forest classifier using 17 mRNA features selected by SVC-L1 to distinguish regional metastatic from primary tumors	mRNA markers accurately predict metastasis status (AUROC 0.95).	Limited feature interpretability; retrospective genomic data.
[29]	0.8732 (Acc)	Primary/Metastatic (SKCM-TCGA)	Logistic Regression model using 15 features to distinguish metastatic tumors from regional/primary tumors	Confirmed multi-omics approach is feasible for identifying metastatic progression markers.	Model performance dropped when classifying distant metastasis (M2).
[30]	0.8947% (Acc)	Primary/Metastatic (SKCM-TCGA)	SVC-W model using selected genomic features to classify metastatic vs. primary tumors	mRNA expression profiles (17 features) showed the highest predictive power for metastasis classification (AUC 0.95, Acc 89.47%).	Methylation features were less predictive than mRNA/miRNA; complex feature selection required.

Design diagrams



Dataset: HAM 10000

Source: Human Against Machine with 10,000 Training Images (ISIC Archive, Vienna & Harvard collaboration)

Size & Variety: 10k+ dermoscopic images across 7 lesion types with diverse skin tones and imaging conditions.

Ground Truth: Expert-verified diagnoses (histopathology/follow-up); segmentation masks available.

Why HAM 10000?

- Standard benchmark for lesion classification
- Rich class diversity, ideal for CNN & CLDM training
- Enables comparability with prior research

Challenge: Cross-sectional (no time-series progression).

Solution: Use diffusion-based generation to create synthetic temporal pairs (6m, 12m, 24m) for prognostic modeling.

Algorithm

Input: Current lesion image I_now
Output: Predicted lesion images at 6m, 12m, 24m (treated & untreated) with uncertainty map

1. **Preprocessing**
   Resize, normalize, and remove artifacts from I_now.
2. **Segmentation (U-Net)**
   Extract lesion mask M from preprocessed image.
3. **Latent Encoding (VAE)**
   Encode I_now into latent space z_now.
4. **Conditioned Diffusion Model**
   - Add noise to z_now → z_noisy.
   - Input (z_noisy, M, horizon token, scenario token) into latent UNet.
   - Denoise iteratively to generate z_future.
5. **Decoding (VAE)**
   Decode z_future back to image space → predicted lesion image.
6. **Uncertainty Estimation**
   Repeat K times with different noise seeds.
   Compute variance across predictions → U_map.
7. **Output**
   - Display multi-horizon forecasts (6m, 12m, 24m).
   - Show treated vs untreated comparisons.
   - Overlay uncertainty map for transparency.

Implementation

The implementation of the proposed Skin Lesion Segmentation and Prognostic Modeling System was carried out using a combination of modern machine learning frameworks, computer vision libraries, and medical imaging utilities. The development pipeline was designed to ensure modularity, scalability, and reproducibility.

Software Environment

The system was implemented in the Python 3.10 programming language due to its extensive ecosystem of scientific and deep learning libraries.

Implementation

Hardware Environment

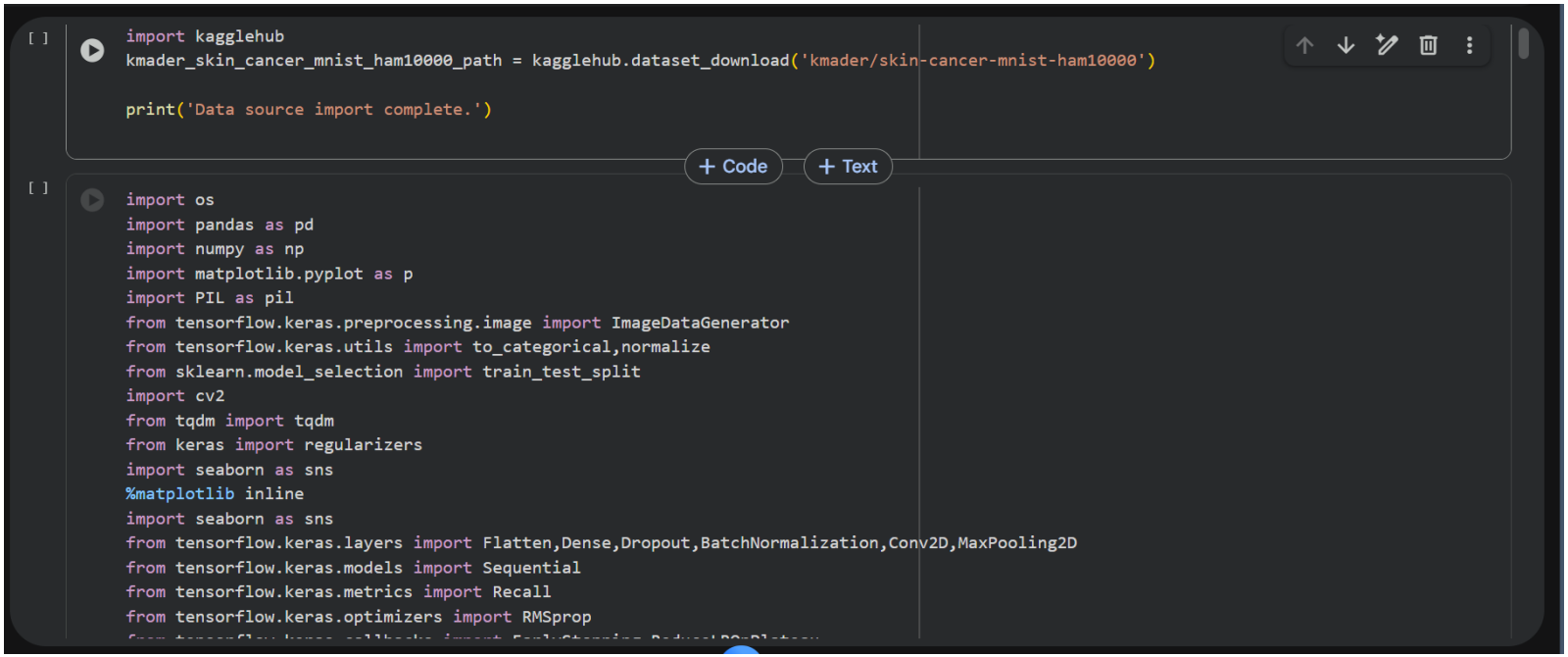
The training and experimentation were conducted on cloud GPUs to leverage

Component	Specification
GPU	NVIDIA Tesla T4 (16 GB VRAM)
CPU	Intel Xeon 2.3 GHz (dual-core)
RAM	16 GB
Storage	100 GB (Kaggle persistent volume)
Runtime	Kaggle GPU Environment

The use of GPU acceleration was essential due to the large dataset size and the computational intensity of both segmentation and diffusion-based prognostic modeling.

Implementation

Screenshot of the GPU runtime environment used for model



```
[ ]
import kagglehub
kmader_skin_cancer_mnist_ham10000_path = kagglehub.dataset_download('kmader/skin-cancer-mnist-ham10000')

print('Data source import complete.')
```

+ Code + Text

```
[ ]
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as p
import PIL as pil
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.utils import to_categorical,normalize
from sklearn.model_selection import train_test_split
import cv2
from tqdm import tqdm
from keras import regularizers
import seaborn as sns
%matplotlib inline
import seaborn as sns
from tensorflow.keras.layers import Flatten,Dense,Dropout,BatchNormalization,Conv2D,MaxPooling2D
from tensorflow.keras.models import Sequential
from tensorflow.keras.metrics import Recall
from tensorflow.keras.optimizers import RMSprop
```

Results

The final model achieved an overall **test accuracy of 72.8%** on the stratified 5k subset.

Metric	Value
Overall Test Accuracy	0.728
Average Precision (macro)	0.38
Average Recall (macro)	0.31
Weighted F1-Score	0.69

Per-Class Performance Summary

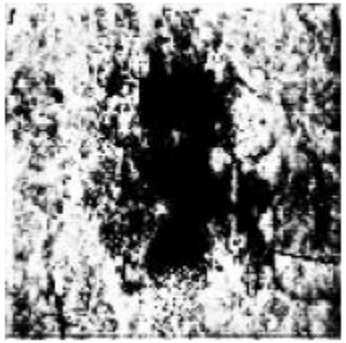
Class	Precision	Recall	F1-Score
akiec	0.55	0.38	0.44
bcc	0.38	0.31	0.34
bkl	0.46	0.38	0.42
df	0.00	0.00	0.00
mel	0.45	0.18	0.26
nv	0.80	0.95	0.87
vasc	0.00	0.00	0.00

Results

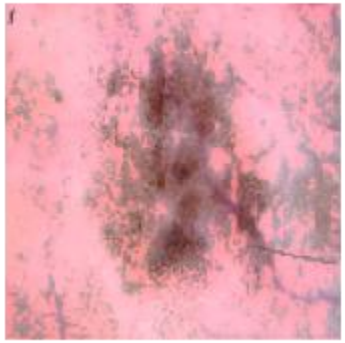
Sample 4 — True: bkl, Pred: nv

Original (bkl)

Predicted Mask

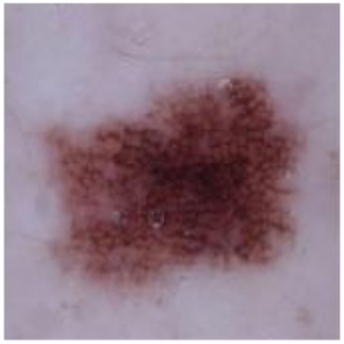


Prediction: nv

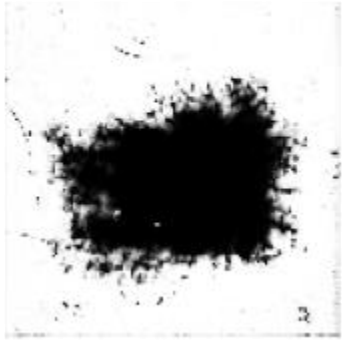


Sample 1 — True: nv, Pred: nv

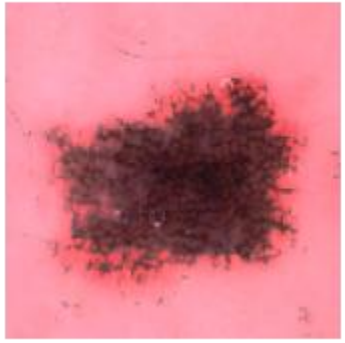
Original (nv)



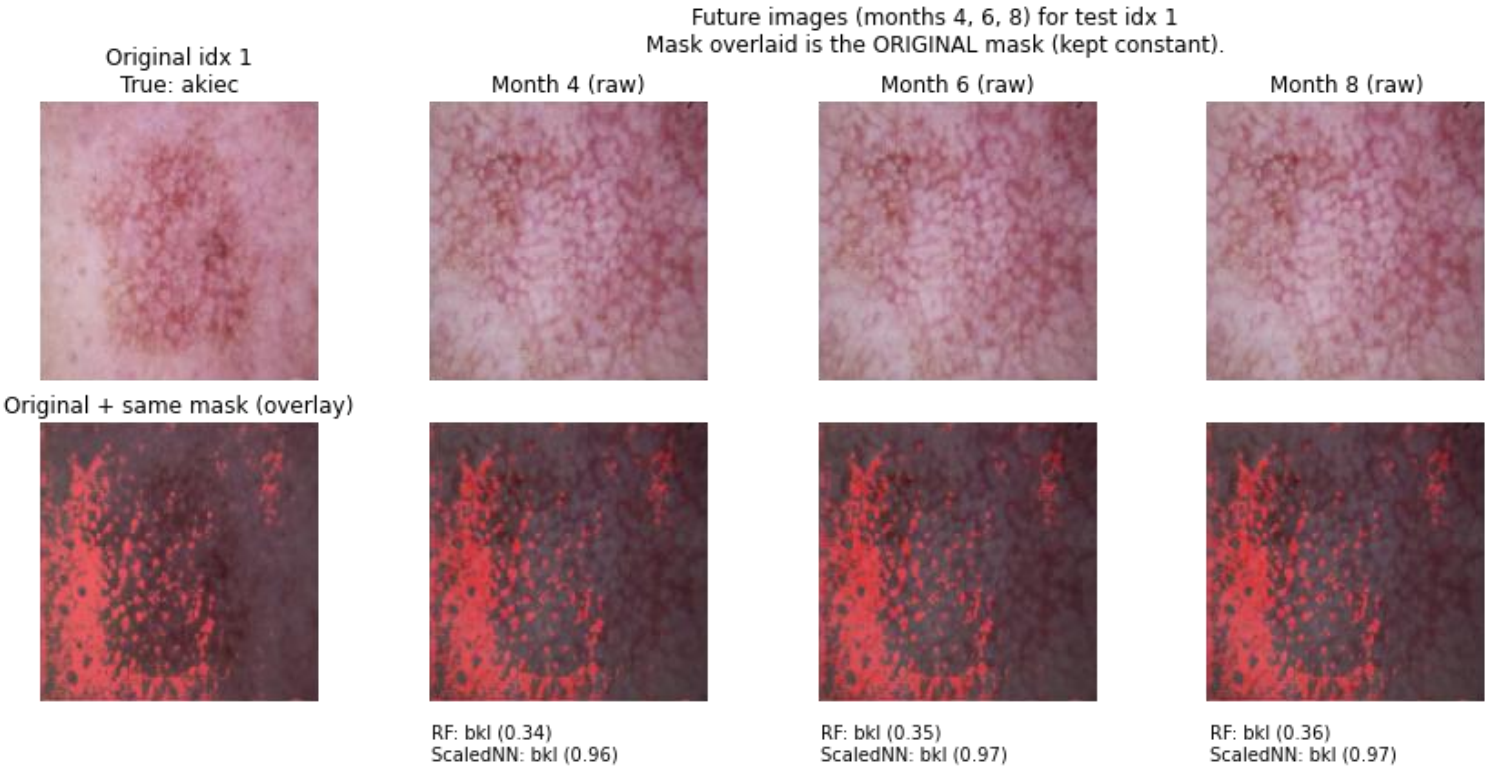
Predicted Mask



Prediction: nv



Results


Future images (months 4, 6, 8) for test idx 1
Mask overlaid is the ORIGINAL mask (kept constant).
Original idx 1
True: akiec
Month 4 (raw)
Month 6 (raw)
Month 8 (raw)
Original + same mask (overlay)
RF: bkl (0.34)
ScaledNN: bkl (0.96)
RF: bkl (0.35)
ScaledNN: bkl (0.97)
RF: bkl (0.36)
ScaledNN: bkl (0.97)

Conclusion

The goal of this project was to design, implement, and validate a new deep learning framework aimed at prognostic forecasting of skin lesions evolution. We addressed an important gap in the current AI diagnostics of skin lesions, which are often limited to static classifiers with predictive labels. The overall framework we developed for prognostic forecasting is an online tool which that is dynamic, multi-horizon, and scenario-based. The distinct contribution of this study is the successful application of a generative diffusion model diagnostic framework with a segmentation backbone to create clinically relevant predictions for multi-faceted classifications.

References

1. C. Metta et al., ”Advancing Dermatological Diagnostics: Interpretable AI for Enhanced Skin Lesion Classification,” Diagnostics, vol. 14, no. 7, Art. no. 753, 2024, doi: 10.3390/diagnostics14070753.
2. P. R. Kshirsagar et al., ”Deep Learning Approaches for Prognosis of Automated Skin Disease,” Life, vol. 12, no. 3, Art. no. 426, 2022, doi: 10.3390/life12030426.
3. S. Q. Gilani and O. Marques, ”Skin lesion analysis using generative adversarial networks: a review,” Multimedia Tools Appl., vol. 82, pp. 30065–30106, 2023, doi: 10.1007/s11042-022-14267-z.
4. N. Codella et al., ”Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC),” arXiv e-prints, arXiv:1902.03368, 2019.
5. C. Metta et al., ”Exemplars and Counterexemplars Explanations for Image Classifiers, Targeting Skin Lesion Labeling,” arXiv e-prints, arXiv:2302.03033, 2023.
6. K. Abhishek, J. Kawahara, and G. Hamarneh, ”Predicting the clinical management of skin lesions using deep learning,” Sci. Rep., 2021, doi: 10.1038/s41598-021-87064-7.